OPTIMAL DYNAMIC PRICING STRATEGIES FOR MOBILE COMMUNICATION NETWORKS

Elena Dimitrova Fitkov-Norris

September 2002

A thesis submitted for the degree of Doctor of Philosophy of the University of London

Department of Electronic and Electrical Engineering

University College London

Torrington Place

London WC1E 7JE

Abstract

Techniques from engineering, economics and control theory are used in this thesis to investigate the effectiveness of dynamic pricing for demand control and capacity optimisation in cellular mobile networks. The scope is extended to include pricing strategies that can provide a certain target revenue for the network operator. Algorithms for the application of dynamic pricing in voice and data networks are suggested. Mathematical models are developed to predict the effect of dynamic pricing on the network operator's market share and the overall user demand, including the effect of variable tariffs on user mobility. The question of setting the optimal tariff for a given system load is addressed and three dynamic price setting methods suggested. The first, competition driven *ad hoc* pricing, is used to identify the most sensitive parameters in the model, namely the revenue generated and the level of call blocking in the network. Two further tariffs (linear revenue attainment and optimal revenue attainment) are then developed for controlling the system and ensuring optimal behaviour. The tariffs are tested using a seven-cell cellular model developed with OPNET™. Simulation results show that the performance of the competition driven *ad hoc* and linear revenue attainment linear pricing strategies is varied and they lead to either a significant reduction in the revenue of the network operator or the welfare of users. The optimal revenue attainment price setting strategy, on the other hand, is shown to be an effective tool for generating the desired revenue, while decreasing the average price in the network and increasing the number of successful calls. In addition, it is suggested that the optimal dynamic pricing strategy could potentially increase a network operator's market share by up to 10% compared to traditional pricing policies, thus offering a viable pricing alternative.

To Ben, Jonathan and my parents

Acknowledgements

I would like to thank my supervisor Dr. Ahmad Khanifar for his patience, understanding and sound advice, my second supervisor Prof. Keith Ward for his help in getting this thesis started all this time ago, and his constructive criticism and support throughout the project. A very special thank you is due to Mrs Nina Thornhill for her timely help with Matlab and extremely useful advice.

I would like to thank my sponsor BT for funding this project and my mentor Katherine Doyle for her constant help and encouragement.

I would also like to thank my colleagues: Jason for his help with Unix programming and web design, Anna and Naomi for their help with the survey, Aikaterini for her help with OPNET™, Giannis for his help with ATM and Chrisina for her help with the model. Thanks also to Charles, Rob, Nishita and Simon for their moral support.

I would also like to thank my husband, Jonathan, for being there when I needed him, and finally, my mother Vassilka Fitkova, for giving it all up and making it possible for me to continue this PhD.

Table of Contents

List of Figures

List of Tables

Glossary of Abbreviations and Acronyms

ALOHA	A transmission system in which a packet is broadcast when ready, and if a collision occurs it is retransmitted. A variation called Slotted ALOHA sends packets at specific time slots to reduce the number of collisions.
AMPS	Advanced Mobile Phone Service (USA)
APC	Automatic Power Control (in CDMA)
ATM	Asynchronous Transfer Mode
AUC	Authentication Centre
BCCH	Broadcast Control Channel
BSC	Base Station Controller
BSS	Base Station Sub-system
BSys	Billing System
BTS	Base Transceiver Station
CDMA	Code Division Multiple Access – a radio interface access protocol specified by IS-95
CDR	Call Detail Record Charging ID
CGF	Charging Gateway Functionality (GPRS networks)
DINAMO	Dynamic revenue optimisation program developed for American Airlines.
Dynamic (real time) pricing	Price for the service varies in time depending on level of demand
EIA	Electronic Industries Association
EIR	Equipment Identity Register
E-TDMA	Enhanced TDMA
ETSI	European Telecommunications Standards Institute
FDMA	Frequency Division Multiple Access – a radio interface access protocol used by analogue cellular systems

Flat rate service	Customers pay a basic fee for connection and service and do not pay for any additional phone calls
GGSN	Gateway GPRS Support Node
GoS	Grade of Service defined as the percentage of lost calls in a telecommunications network.
GPRS	General Packet Radio Service
GPS	Global Positioning System
GSM	Global System for Mobile Communications
HLR	Home Locations Register
IMSI	International Mobile Subscriber Identity
IMT-2000	International Mobile Telecommunications by the year 2000 standard
ISP	Internet Service Provider
Measured service	Customers pay a basic fee for connection and for all additional phone calls they make
MAC	Medium Access Control Protocol
MS	Mobile Stations
MSC	Mobile Switching Centre
MTSO	Mobile Telephone Switching Office (AMPS networks)
NMT	Nordic Mobile Telephone
NSS	Network Switching Sub-system
OFTEL	Office of Telecommunications
OMC	Operations and Management Centre
OSI	Open Systems Interconnection
OSS	Operation Sub-System
PBCCH	Packet Broadcast Control Channel
PDP	Packet Data Protocol
QoS	Quality of Service, defined as the overall voice and transmission quality in a telecommunications network.

RACH	Random Access Channel used for access to the GSM network based on the slotted ALOHA method.
Real time pricing	see *dynamic pricing*
Repression	Corresponds to the reduction in customer usage due to tariff changes
RPS	Revenue per subscriber
SDH	Synchronous Digital Hierarchy (fixed networks)
SDU	Service Data Unit
SGSN	Serving GPRS Support Node
SIM	Subscriber Identity Module
TACS	British Total Access Communication System
TIA	Telecommunication Industry Association
TDMA	Time Division Multiple Access – a radio interface access protocol used by GSM networks
TRAU	Transcoder and Rate Adapter Unit
TSC	Transit Switching Centre
VLR	Visitors Location Register
UMTS	Universal Mobile Telecommunication Systems
VFM	Value For Money
WAP	Wireless Application Protocol

Chapter 1

1.1 Introduction.

Technological developments in electronics in the last 20 years have transformed the lives of ordinary people and redefined the way business is done. The world-wide web, e-commerce and dot.com companies have become household names, while at the forefront of this digital revolution, cellular network operators are providing wireless applications which allow access "anytime, anywhere". In fact, there are suggestions that in the very near future mobile and wireless terminals will completely replace fixed terminals[1]. The predicted growth in demand for future network services is phenomenal as can be seen in Figure 1-1.

The UK mobile cellular telephony market, in particular, has expanded very rapidly in the past decade. At the end of the first quarter of 1999, mobile phone penetration was almost 30% - up by 5% from the end of the previous year[2]. The revenue and call minutes generated by users have also increased in proportion to the increase in subscribers (see Figure 1-2 and Figure 1-3 below) and, by 2003 cellular revenue is expected to grow by a further 200% [1], with data playing an increasingly large part [2].

[1] Kenny Hirschhorn, Group Director of Strategy, Imagineering and Futurology with Orange, claims that by 2005 all communications will be performed using a small, voice activated device that can be worn as an ear-ring [2].

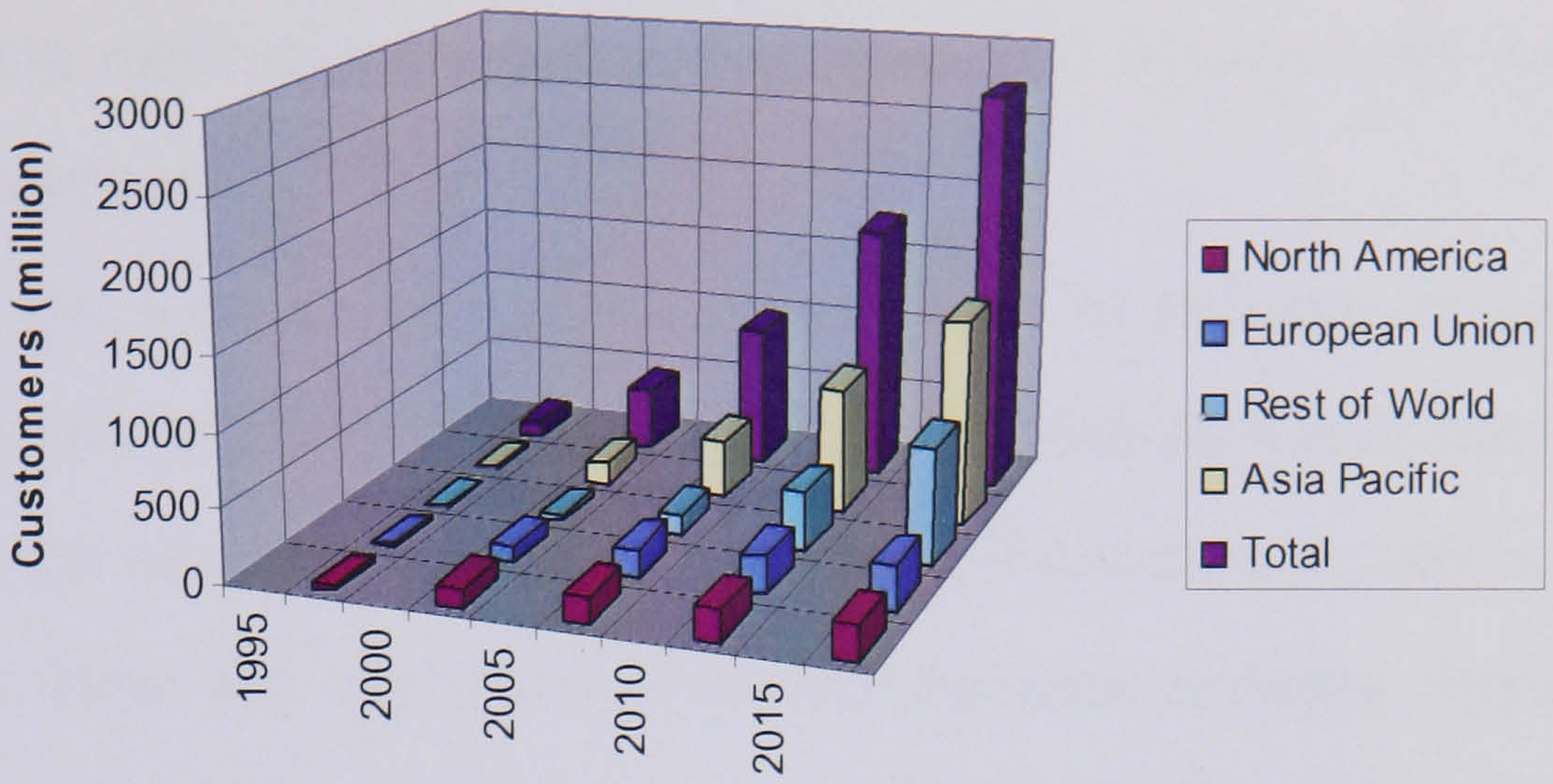

Figure 1-1 *Long term forecast world-wide mobile market (Feb. 1999) [3]*

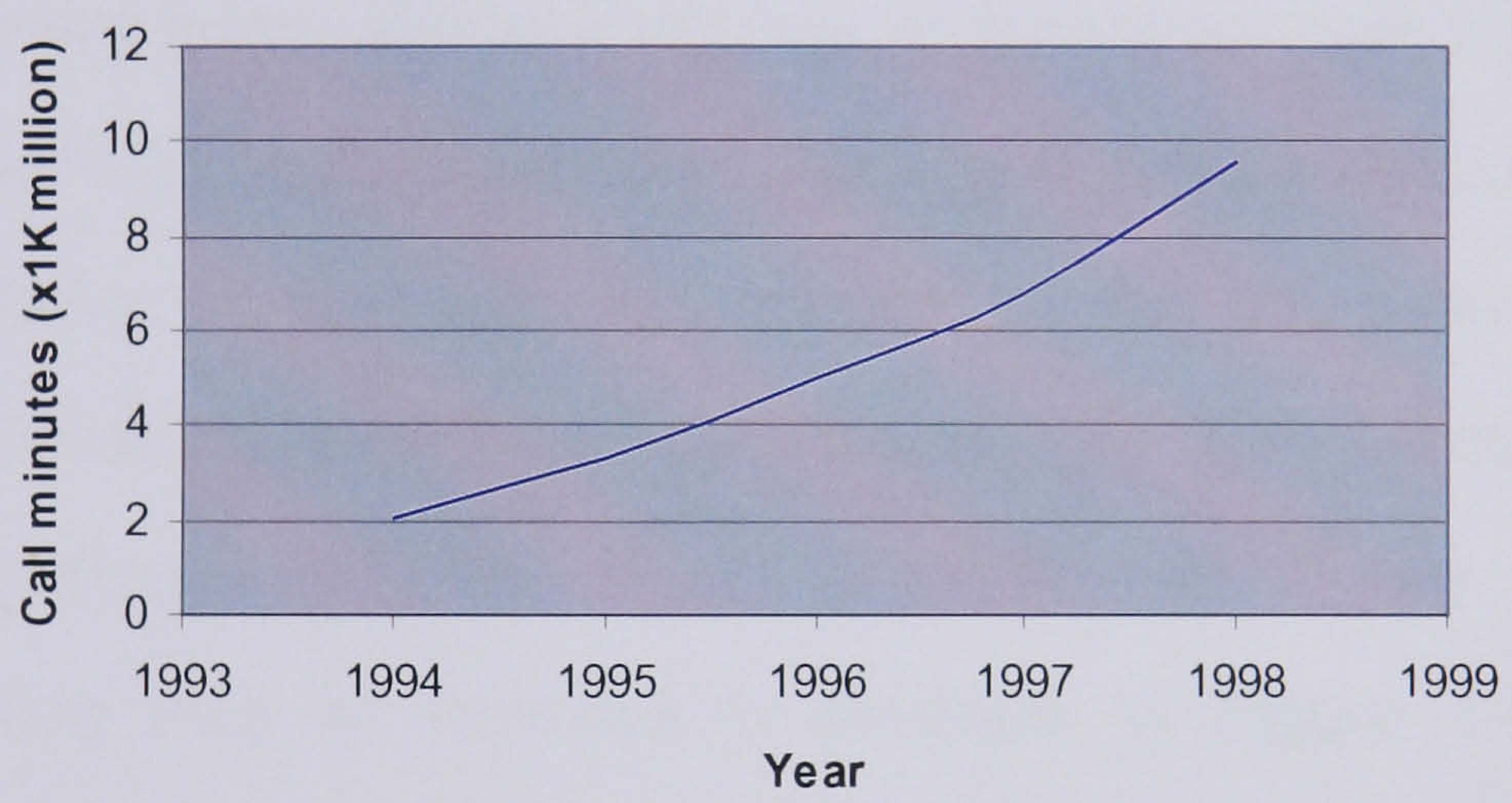

Figure 1-2 *Number of call minutes in the UK market (April 1999) [4]*

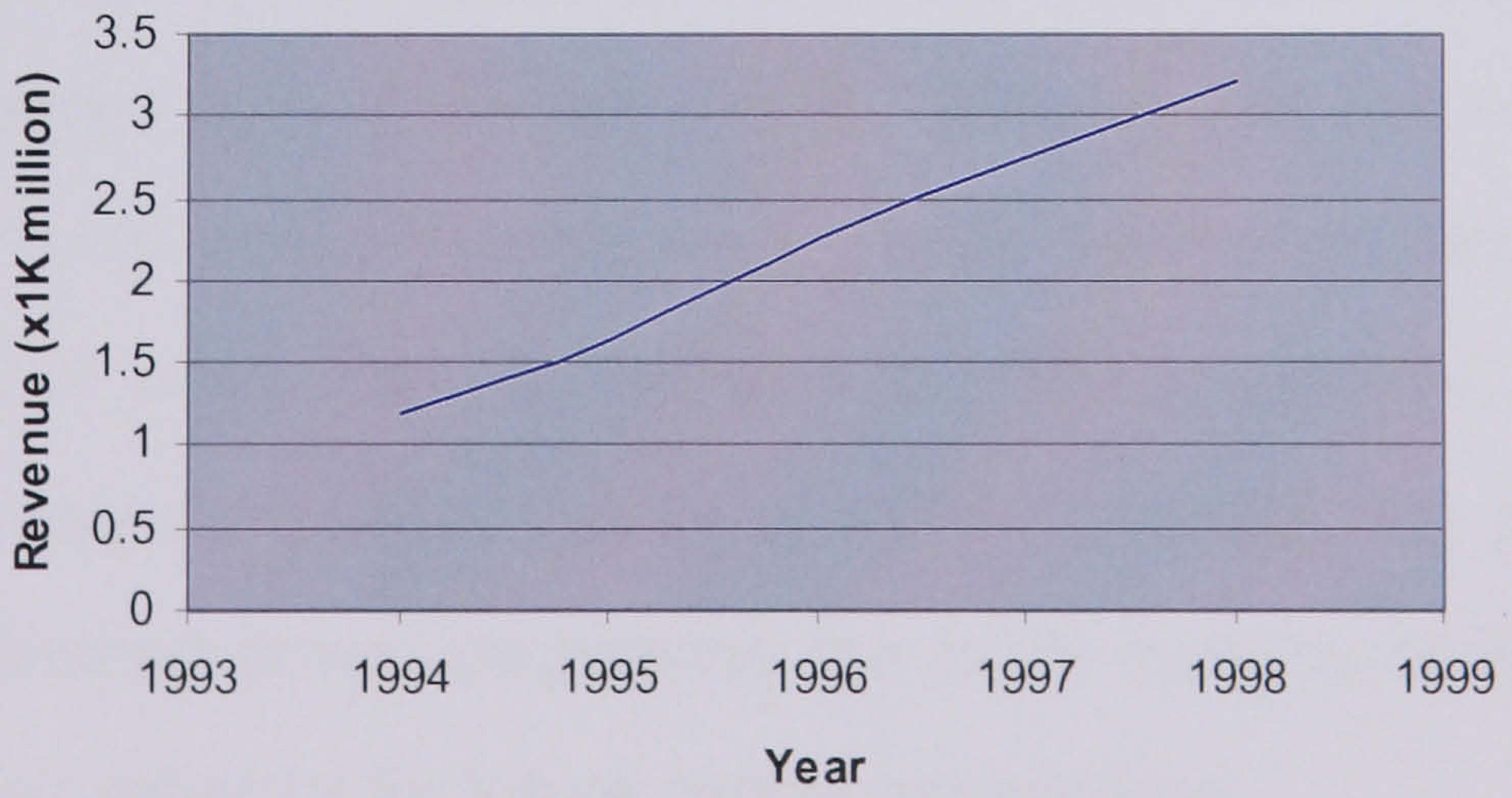

Figure 1-3 *Revenue generated from cellular networks (April 1999) [4]*

The increase in demand for network services has led to substantial investment in the network infrastructure and one UK operator, for example, invests £10 million a week into their network [5].

However, the total network capacity available to network operators is limited, not only by the amount of capital they can afford to invest but also by the total bandwidth allocated to cellular telephony. Furthermore, technological limitations also hinder the expansion of cellular network capacity. Typically, a GSM network can supply bandwidth up to 2Mbit/s/km^2, while an optical Synchronous Digital Hierarchy (SDH) fixed line network can deliver up to several hundred Mbit/s/km^2 and as demand grows its capacity can easily be increased many times with relatively low incremental costs [2]. This is the chief limiting factor limiting the expansion capacity of cellular networks.

In addition, a defining characteristic of the UK cellular market is the aggressive rivalry between service providers [6]. This fierce competition has let to a significant reduction in call charges (see Figure 1-4). In fact, the relatively slow rate of increase in revenue in Figure 1-3 is due to the decreasing prices, partially obviating the effect of the rapid growth in mobile phone usage.

Therefore, network operators today are faced by a very real and difficult problem caused by the overall rise in demand. On the one hand, further increases in user demand would require installation of additional infrastructure, a costly option, which can only be used to a limited extent. On the other hand, due to the very competitive nature of cellular markets, the prices of services are being pushed down, decreasing the profit margins of the operators and reducing their potential for future capital expenditure.

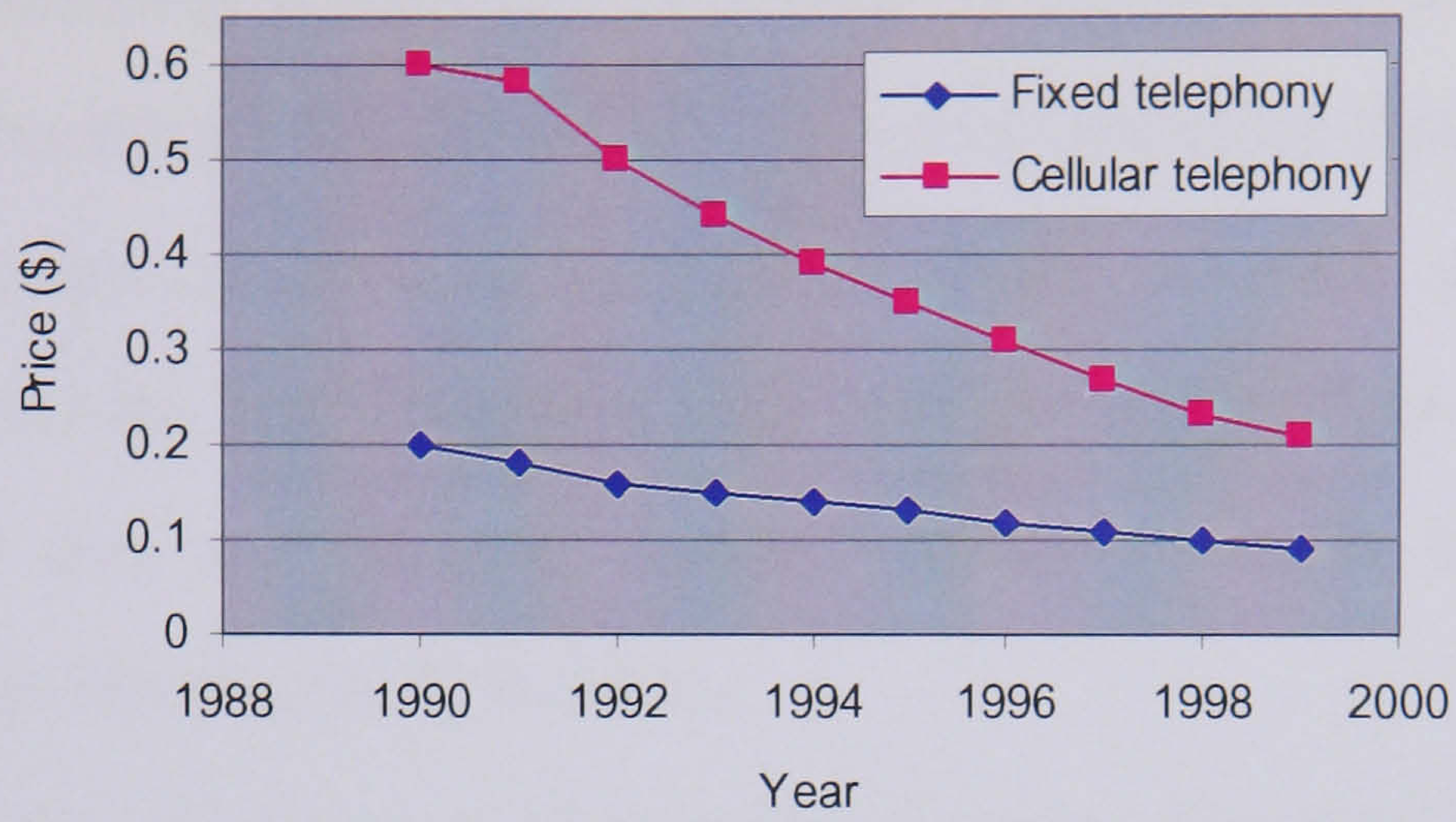

Figure 1-4 *Comparison of average cost per minute for fixed and cellular telephony in Europe*[3]

To combat the adverse effects of the decreasing profit margins, the principle aim of network operators is to optimise their return on invested capital by maximising the return per subscriber and use of available network capacity. This task is complicated by the fluctuation of user demand for telecommunication services according to time of day. The difference between peak and off-peak demand is very significant, with a magnitude of difference of up to 20 to 1 between the busy and quiet periods [7]. An additional dimension to the problem in cellular networks is the mobility of the users, who can move freely around the network, making the accurate prediction of expected demand harder. Once put into place, the available capacity in individual cells is fixed and although engineering solutions, such as dynamic channel assignment, alternate routing and dynamic cell sizing, have been suggested for moving capacity around the network, these are at the cost of increased signalling overheads and/or interference. In addition, these strategies will fail if all cells in the network are busy at the same time.

[3] France, Germany, Italy, the Netherlands, Sweden and the UK.

An alternative approach to shift demand for optimising the utilisation of network resources is via the use of pricing. In order to smooth out fluctuations in demand, this thesis suggests that the pricing could be dynamic; the price for calls will change as demand fluctuates. It will rise with demand, deterring some new entrants from choosing to access the network, and *vice versa*. This is not a new idea and dynamic pricing has, for example, been implemented successfully in the electricity industry.

This thesis will address all aspects of pricing faced by network operators starting with the economic objectives of pricing, through the pricing options and finally, several dynamic pricing strategies for setting of dynamic pricing and evaluating their optimality are suggested.

This is a multidisciplinary subject, which combines techniques from operational research, marketing, economics, control theory and engineering into a comprehensive study of pricing for the mobile telephony market. This complex task is further complicated by the fact that accurate forecasting of demand for new telecommunication services is difficult. This is due to the exceptionally fast growth of the market, substitution effects that are difficult to model, such as the interchange in usage of fixed and mobile terminals (Kelly [8] and Boden [9]) and a lack of accurate information. Furthermore, any predictions made about a particular market cannot always be extended to another market. It should be borne in mind, therefore, that the conclusions drawn in this thesis are derived from data gathered in the UK. As a result they would be applicable to the UK market and should only be extended to other markets with caution.

1.2 Structure of the Thesis.

Chapter 2 presents the problem of limited network capacity from an engineering point of view. It begins with an outline of the structural and operational characteristics of current GSM and GPRS systems as well as the future IMT-2000 systems. The factors affecting the capacity and spectral efficiency of the systems mentioned above are identified, followed by definitions of terms commonly used in traffic theory for the measurement of network capacity utilisation. The temporal distribution of cellular traffic is presented and followed by techniques suggested for better utilisation of the available network resources in 2nd and 3rd generation systems. These include dynamic channel assignment, alternate routing and dynamic cell sizing. The techniques will be described and their advantages will be weighed against their drawbacks. An alternative means for fitting demand more closely to the available network capacity through dynamic pricing is suggested and its advantages discussed. Finally, a significant economic implication of the application of dynamic pricing is identified and a link between dynamic pricing and control theory explored.

Chapter 3 focuses on pricing from an economic point of view. The economic objectives of network operators will be identified, followed by an overview of the tariffs currently used by network operators for meeting these objectives. A mathematical model intended to predict the effect of dynamic pricing on the market share of network operators is presented. The choices faced by network operators when determining the tariff (charging units) for accounting of resource usage by customers are presented, before describing some practical implementations of dynamic pricing. Dynamic pricing algorithms are presented for voice and packet based data networks[4] before engineering issues arising from the implementation of dynamic pricing, such as the signalling overhead and the best position for the algorithm are discussd. Finally, possible complications due to the interconnectivity of cellular networks with the fixed network are examined.

Chapter 4 continues the discussion of the expected effect of dynamic charging from an economic point of view, in particular its effect on user behaviour. A comprehensive mathematical model for evaluating this effect is introduced. This takes into account the price elasticity of user demand as well as the existing pricing bias in the network and a substitution effect due to the fixed network. In addition, a user mobility model showing the effect of price on user mobility is also introduced. The discussion then focuses on the final stage in any pricing policy: the setting of monetary units to the charging units. A literature overview of suggested optimal price setting strategies from an economic and consumer welfare point of view will be given and the drawbacks of each approach highlighted and an alternative "provider-oriented" approach evaluated.

Chapter 5 presents a simple, market-driven, *ad-hoc* dynamic pricing setting policy, which takes into account the minimum and maximum price the network providers want, or have to, charge and takes the current market price as the average price in the network. A seven-cell OPNET™ simulation model developed to test the effect of dynamic pricing is described. Simulation results with the *ad hoc* pricing policy indicate that the behaviour of the network is very sensitive to the shape of the proposed pricing function.

[4] Although a dynamic pricing algorithm for packet-based networks is suggested, due to limited resources the optimal dynamic pricing strategies will be implemented only for voice based services.

Chapter 6 introduces a feedback control theory approach to the setting of dynamic pricing. The revenue generated by the network operator and the percentage of blocked calls are identified as the most sensitive variable in the system and, therefore, recognised as very suitable controlled variables. The mathematical framework of an alternative revenue attainment approach to dynamic pricing is, therefore, given, guaranteeing a predefined minimum revenue for the network provider and assuming dynamic price linearity. Simulation results showing the effectiveness of the revenue attainment model are presented and discussed. Finally, the requirement for the linearity of the pricing function is dropped and an optimal shape found, through modification of the revenue attainment model, using calculus of variations and control theory. A comprehensive model for the determination of optimal dynamic price setting strategies for any type of demand assumptions is presented and simulation results reported.

In chapter 7 the expected effect of the ad hoc, the linear revenue attainment and optimal revenue attainment pricing strategies on network operators' market share is calculated, with the aim of identifying the optimal pricing for maximisation of the total user database.

Chapter 8 summarises the conclusions of this work and outlines suggestions for further research and development.

Chapter 2

This chapter will begin with a brief historic overview of the development of cellular networks, followed by an outline of the structural and operational characteristics of current GSM and GPRS systems as well as the future IMT-2000 systems. The factors affecting the capacity and spectral efficiency of the systems mentioned above will be identified, followed by definitions of terms used in traffic theory for the measurement of network utilisation. The temporal distribution of cellular traffic will be presented, followed by techniques suggested for better utilisation of the available network resources in 2nd and 3rd generation systems. Dynamic channel assignment, cell sizing and alternate routing will be described and their drawbacks acknowledged. An alternative means for fitting demand to the available network capacity through dynamic pricing will then be suggested, followed by a discussion of its advantages. Finally, a significant economic implication at the application of dynamic pricing will be identified and some control theory terminology introduced.

2.1 Mobile Cellular Systems.

2.1.1 An overview.

The first commercial mobile telephone services were established in 1946, and these early systems used a single transmitter to cover an area of 40 - 50 miles. However, this approach offered a very limited system capacity of very few radio channels, which were quickly saturated. For example, in 1970 the Bell system in New York City could support just 12 simultaneous mobile conversations. The 13th caller was blocked [10]. The key in the development of wireless mobile networks was the discovery, in 1947, by engineers in Bell Laboratories that a reduction in transmitter power reduces the signal coverage area, allowing re-use of the available radio frequencies and thus increasing the capacity of the system. This process is called cell splitting and the area covered by a single low power transmitter is called a *cell* (see Figure 2-1). This technique is still being utilised by service operators and, for example, urban areas have more cells per square kilometre than rural ones.

The minimum set of cells using all the available frequencies is called a *cluster*. In first and second generation mobile systems, frequencies cannot be re-used in adjacent cells because of problems with co-channel interference, but clusters can be repeated throughout the network following a pre-set frequency re-use pattern [11].

Figure 2-1 *Cell splitting*

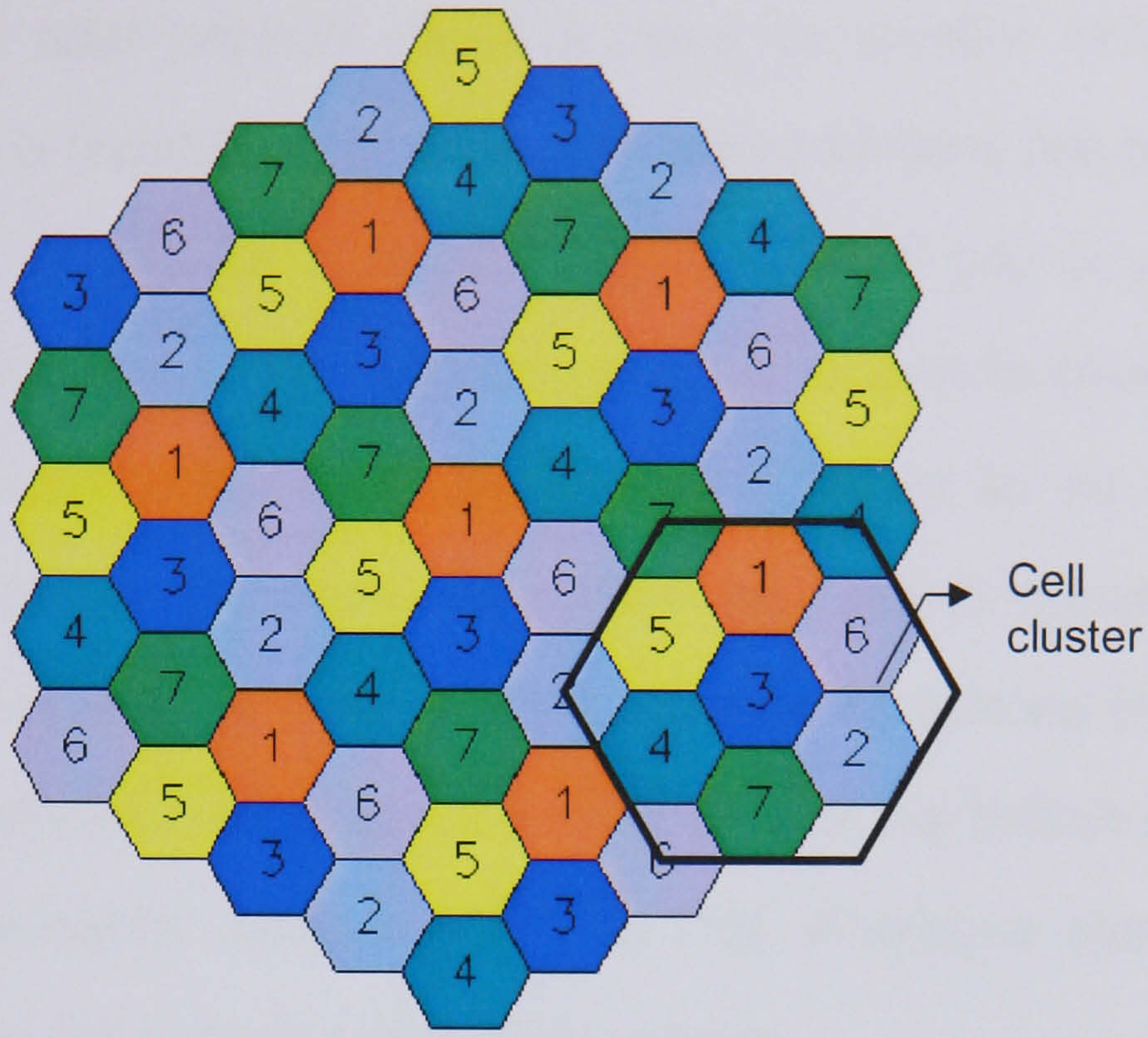

Figure 2-2 *Frequency re-use pattern*

The division of the coverage area into many cells imposes an additional problem with users moving between cells. The network operators have to offer a seamless service, and therefore, as users move to new cells, channels have to be provided for mobile calls to be transferred from one cell to the next. This process is called “handover”. As the users cross cell boundaries, their mobile stations have to “sign off“ from the controller in the old cell and re-register with the controller managing the new cell (see Figure 2-3).

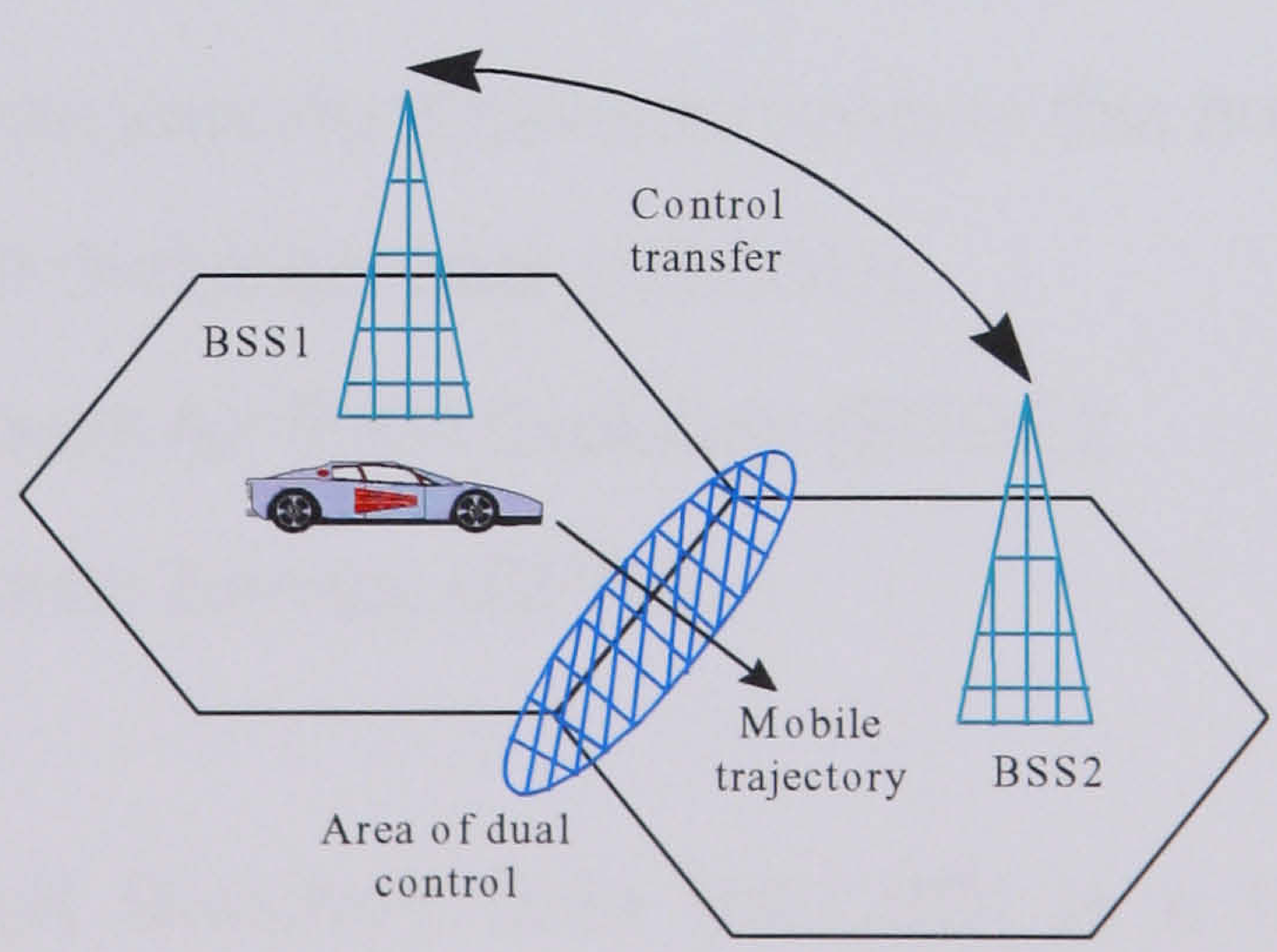

Figure 2-3 *Handover procedure*

The increase in total network capacity, due to smaller cells, turned cellular telephony into a feasible and useful alternative to fixed line telephony. The first generation of mobile systems was designed primarily for the transmission of speech signals, although they were also able to transmit data at relatively low bit rates [12]. They are usually referred to as analogue systems as the radio transmitter does not digitise the speech signals prior to transmission. These systems include the Nordic Mobile Telephone (NMT), the American Advanced Mobile Phone Service (AMPS) and the British Extended Total Access Communication System (ETACS) [12]. Analogue systems are being gradually phased out in favour of digital systems.

Second-generation cellular radio systems are digital and offer higher speech quality, increased capacity and security, and some standards also offer international roaming, *i.e.* mobile stations can connect and use compatible networks abroad. Examples of second-generation systems are the Global System for Mobile Communication (GSM) and DCS-1800 in Europe, and Digital APMS (D-AMPS, IS-54) and PCS-1900 (IS-136, an upgraded specification of IS-54) in the USA.

Enhancements of the GSM and IS-136 second-generation systems are currently being rolled out to enable current networks to support higher rate data services. The most important developments in this area are [13]:

1. High Speed Circuit Switched Data (HSCSD);
2. Enhanced Data Rates for GSM Evolution (EDGE);
3. General Packet Radio Service (GPRS).

1. High Speed Circuit Switched Data (HSCSD) is a GSM bearer service intended to use multiple time slots for increased data rate over the GSM air interface, offering data rates of up to 57.6 Kbit/s. HSCSD may be used for

a wide variety of tele-services and its advantage is that it requires no restructuring of the available GSM infrastructure. Its disadvantage is that it reduces the number of users the cells can support thus effectively reducing cell capacity.

2. EDGE, on the other hand, will use new modulation and channel coding techniques to evolve data services over GSM networks by using as much of the existing physical layer as possible. The disadvantage of EDGE is that it requires new infrastructure installation in the Base Transceiver Stations (BTS) (see section 2.1.2.1 below) and in addition reduces the cell radius because of higher interference. EDGE can support data rates from 22.8 Kb/s to 69.2 Kb/s depending on the channel-coding scheme [13].
3. GPRS uses a packet based approach and will be discussed in greater detail in section 2.1.2.2. The EDGE concept has also been considered for GPRS (EGPRS) and Circuit Switched Data (ECSD).

Third generation of mobile telecommunications standards is now under development, such as the International Mobile Telecommunications by the year 2000 standard (IMT-2000). This standard is known in Europe as the Universal Mobile Telecommunication Systems (UMTS). The operational objective of the new systems is to provide seamless services across various radio and fixed cable environments and under different operational conditions. The specifications include the provision of broadband multimedia (voice, data, video) and a home environment which can be defined by the user, so that consumers experience the same services wherever they are [14][5]. It is intended that migration from second to third generation systems will be

[5] Their development is well on its way and the European Parliament required its member states to put in place a harmonised system for authorising UMTS systems by 1st January 2000 in order to allow the provision of UMTS services by 1st January 2002.

progressive and will be perceived by existing customers as a service evolution which is beneficial, attractive and natural [15].

In the following sections, before identifying the parameters affecting the capacity of the three types of systems outlined above, the architecture of cellular networks will be examined in detail. This would enable us to fully appreciate the serious network management problems faced by cellular operators brought about by limited network capacity. Tools developed by teletraffic engineering theory to measure the effective utilisation of available system capacity and overall system performance will be described, followed by an overview of techniques for offering flexible capacity.

2.1.2 Overview of GSM, GPRS and UMTS Network Architecture.

The basic architecture of the three types of cellular networks studied in this section is very similar. In general, cellular networks consist of Mobile Stations (MS) carried by the user, which can move freely around the network and the fixed support network infrastructure. The MS uses radio frequencies to communicate with the fixed support network. The support network is layered and the degree of stratification depends on user density. In the following section, the principal components and their main functions for the GSM support network are explained in detail and the differences in the structures of the other two networks highlighted.

2.1.2.1 GSM network architecture.

The GSM fixed support network consists of a Base Station Sub-System (BSS), a Network and Switching Sub-System (NSS) and an Operation Sub-System (OSS) (see Figure 2-4).

Each cluster of cells in the network is controlled by a Base Station Sub-system, which consists of two parts: Base Transceiver Stations (BTS) and a Base Station Controller (BSC). The BTS are located in cells and communicate with the MS in the cell; they are responsible for the transmission, reception and management of the radio interface. The BSC, on the other hand, is responsible for the allocation and release of radio channels and handover management within the cell cluster.

The main task of the Network Switching Sub-system (NSS) is to manage the databases needed for subscriber data and to direct the traffic around the network. It consists of Mobile Switching Centres (MSC) and two databases: the Home Locations Register (HLR) and the Visitors Location Register (VLR).

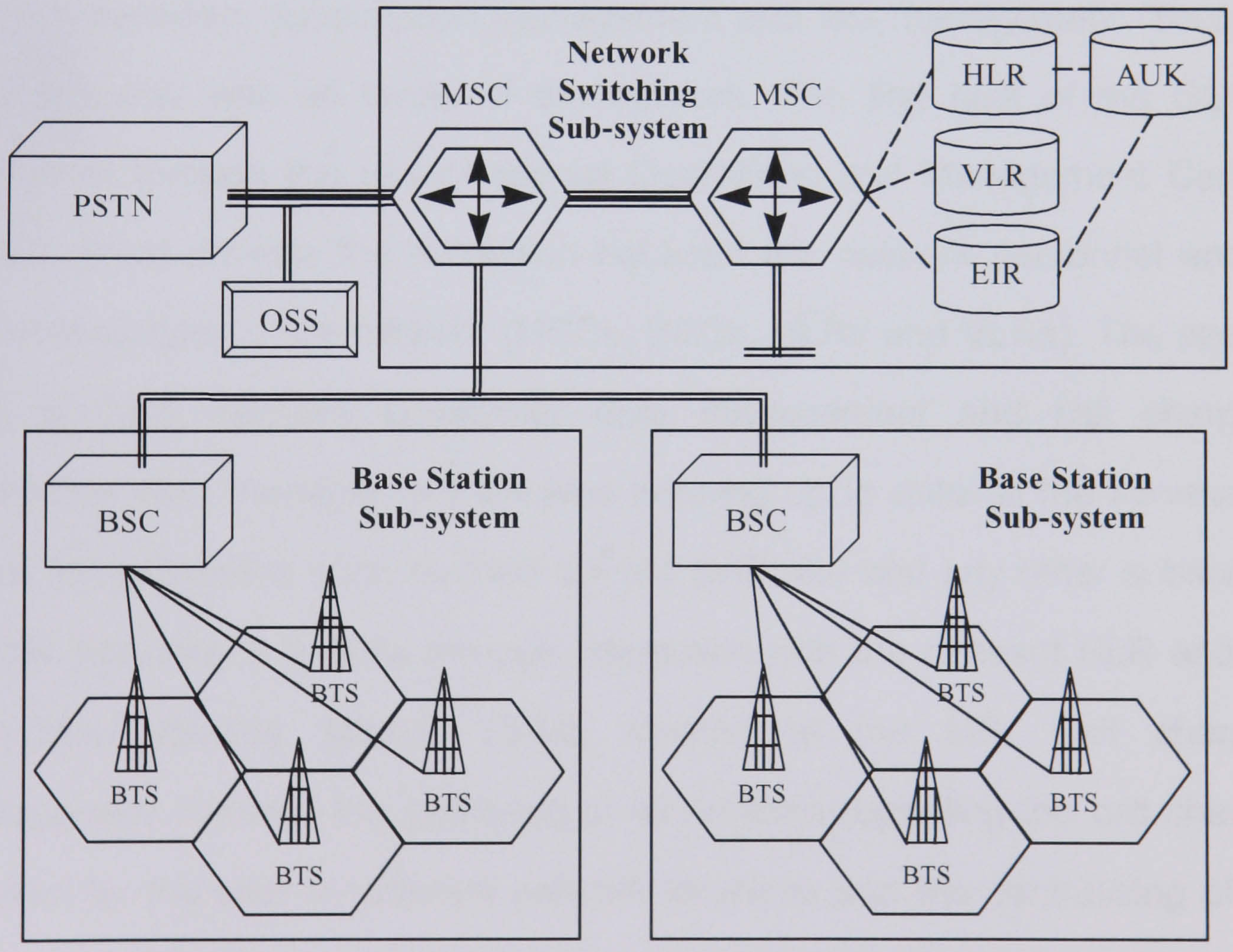

Figure 2-4 *Structure of the GSM mobile system*

Each MSC is responsible for the control of a few BSCs and the set-up and co-ordination of calls to and from GSM users. As the number of users increases, full intermeshing of the MSCs becomes very costly and complicated and so a second layer of Transit Switching Centres (TSC), has to be added to the hierarchy. These are now present in almost all mobile networks [16]. The HLR and VLR associated with the MSC contain details of the physical location of each MS in the area - those registered in the cell and those visiting the cell. For security reasons, an Authentication Centre (AUC) and Equipment Identity Register (EIR) are introduced to check the status of the user attempting to use the network [17].

The main tasks of the Operation Sub-System involve network operation and maintenance, subscription management and MS management. Thus the OSS interacts with all levels of the network. The first task of the OSS is completed through the use of special Operations and Management Centres (OMC) which provide the mediation between the network personnel and all machine entities in the network (MSCs, BSCs, HLRs and VLRs). The second task of OSS involves subscriber data management and call charging. Subscriber data management involves keeping up to date all the information about the subscriber such as their current tariff rate and any other subscriber specific features. It is done through interaction with the relevant HLR and the Subscriber Identity Module (SIM), located at the MS. Call charging management involves the gathering of information regarding the call charges incurred by the user in different network locations and the centralising of the billing data. The final task of the OSS is to update the MS equipment register, which is particularly useful when searching for stolen or misbehaving MSs.

2.1.2.2 GPRS network architecture.

As mentioned earlier, GPRS (General Packet Radio Service) represents an enhancement of GSM that enables it to support packet switched services[6]. As a result, two additional nodes are added for handling packets and the existing BSSs are modified. The two additional nodes are:

- Serving GPRS Support Node (SGSN)
- Gateway GPRS Support Node (GGSN)

The architecture of a typical GPRS network can be seen in Figure 2-5 [18].

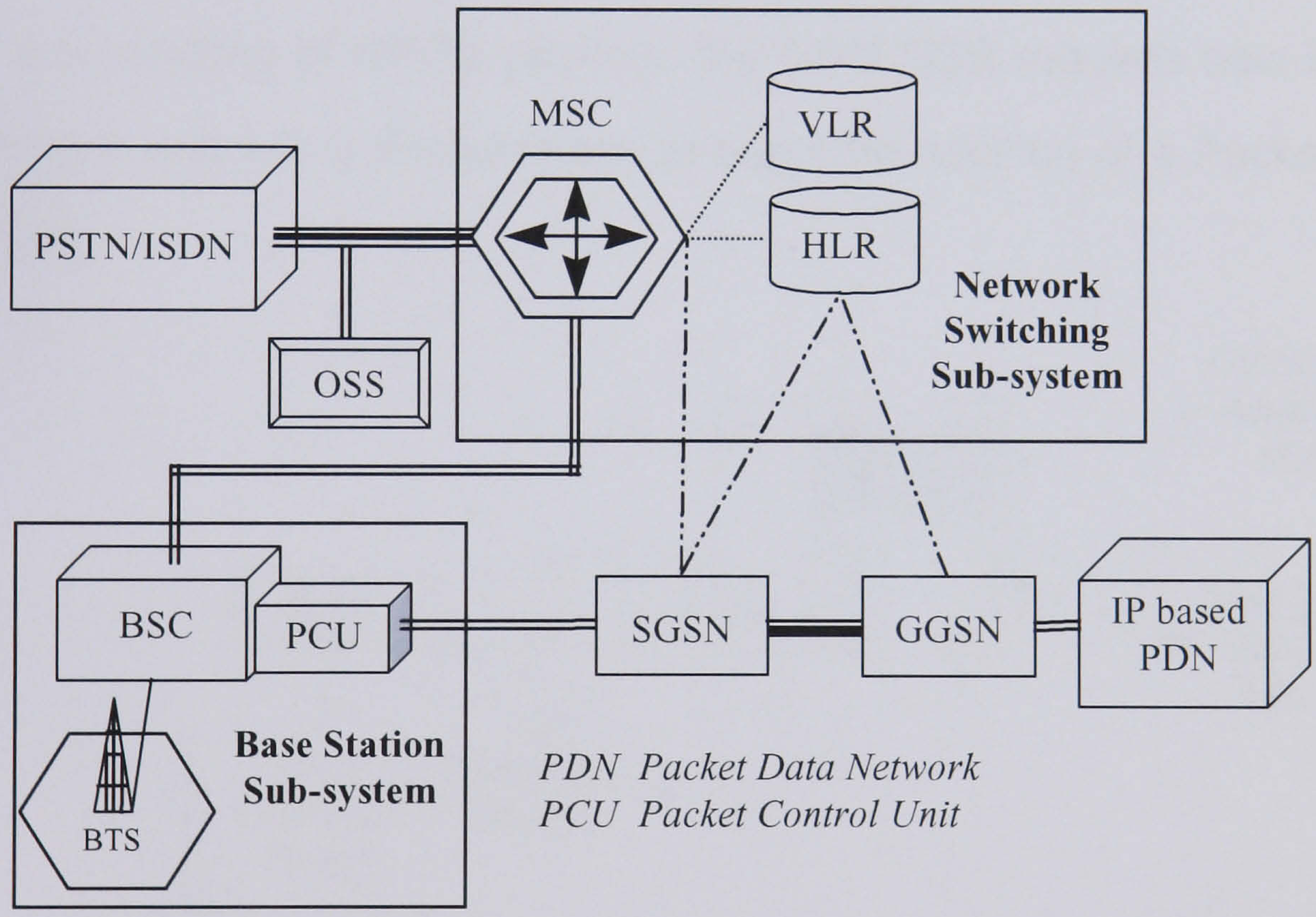

Figure 2-5 *GPRS network architecture*

SGSN sits on the same hierarchical level as the MSC in GSM networks and performs the following functions:

[6] Circuit switched networks, such as GSM, require a dedicated communication path to be established between the initiator of the call and the recipient. There is inherent inefficiency in this approach, which requires the allocation of resources for the entire length of the connection. In contrast, in packet-switched networks, no path is dedicated and the available resources are allocated as each packet arrives depending on demand [18].

1) Packet routing to and from any GPRS-enabled MS in its service area;
2) Session and mobility management;
3) Authentication and ciphering;
4) Packet counting for billing purposes.

GGSN, on the other hand, performs the following functions:

1) Encapsulation of Packet Data Protocol (PDP) packets into GPRS packets;
2) Routing (tunnelling) of Packet Data Units (PDUs) to the serving SGSN;
3) Packet counting for billing purposes.

The GPRS infrastructure utilises the same HLR and VLR databases as the GSM network and also shares the same radio interface resources. To enable the handling of GPRS packets, the GSM BSS requires new hardware and software and this is implemented through the addition of a Packet Control Unit (PCU).

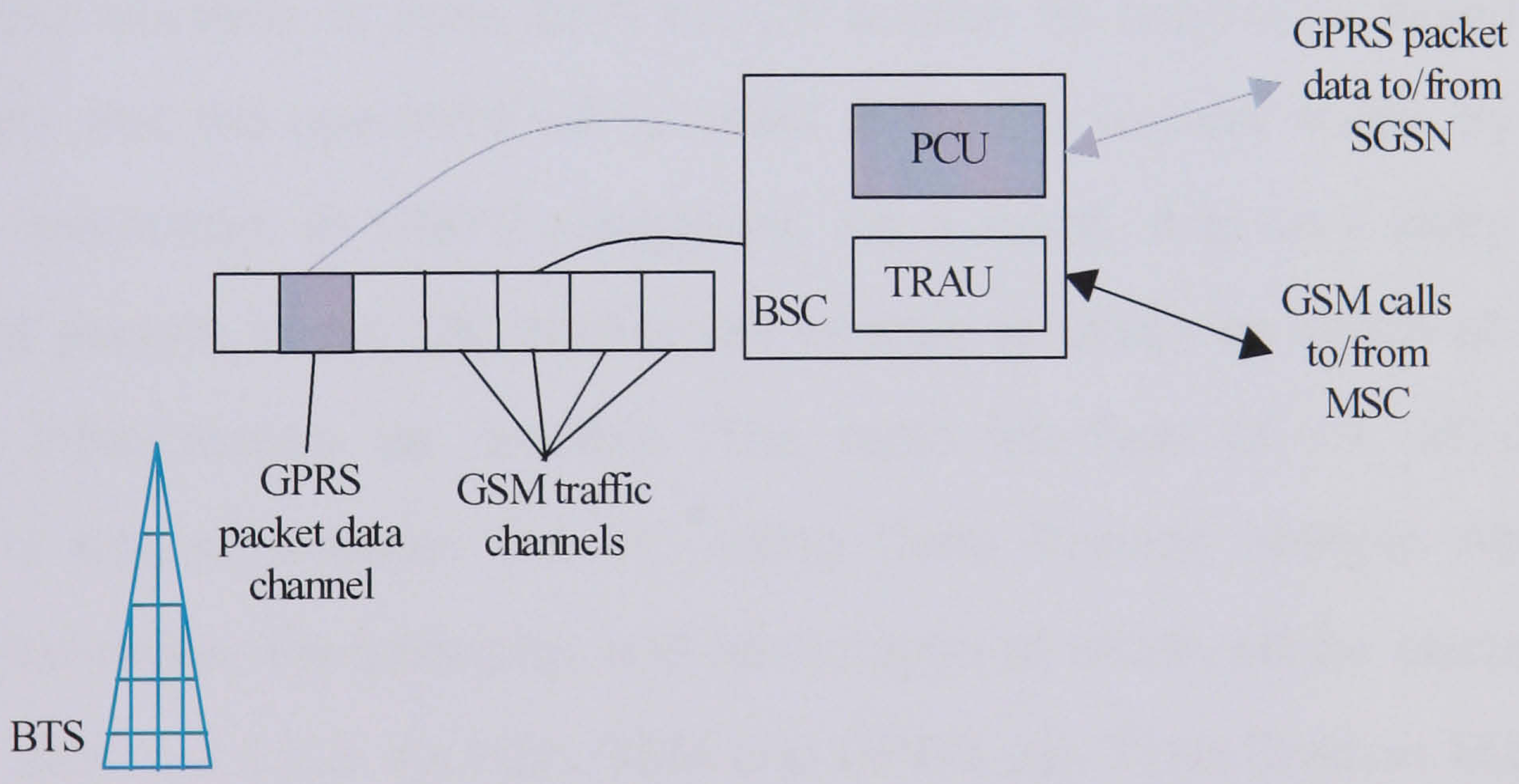

Figure 2-6 *Functions of PCU*

The PCU is responsible for the allocation of GPRS channels between all GPRS enabled MS that wish to transmit at any given time. This process is known as Medium Access Control (MAC) and is performed in all packet-based networks. The PCU also manages the transfer of user data between the MS and the SGSN. The PCU "listens" on the logical radio channels allocated to GPRS enabled MS and forwards the packets as they arrive. The PCU can be implemented in the BTS, BSC or SGSN, although most vendors have chosen to implement it in the BSC.

The European Telecommunications Standards Institute (ETSI) suggests in its GPRS technical specifications [19] that the allocation of GPRS resources is done dynamically, depending on demand. This "capacity on demand" principle aims to ensure that the GPRS resources are optimally utilised.

2.1.2.3 Structure of IMT-2000 (UMTS) network.

The technical specification of the IMT-2000 system allows network operators a choice regarding the network architecture. It has been suggested, however, due to the massive investment of mobile operators in the UMTS licences (for example, in June 2000 the UK auction for bandwidth raised $34 billion [20]), that the operators will, at least at the introductory stage, try and minimise investment in UMTS equipment. As a result, it is very likely that incumbent players in the UK market will choose to utilise as much of their available infrastructure as possible. The radio interface of an IMT-2000 network is shared between the MS using Code Division Multiple Access (CDMA) technique, the principles and advantages of which will be discussed briefly in section 2.1.3.2. As both GSM and GPRS use Time Division Multiple Access (TDMA) technique in their radio access (see 2.1.3.1), the infrastructure for interaction with MSs over the radio interface cannot be reused. However, the rest of the fixed support network can be shared between GSM, GPRS and UMTS users, as indicated in Figure 2-7.

Radio access for the UMTS network is mediated by the Radio Network Sub-system (RNS). The RNS offers the allocation and release of resources to establish connection between the MSs and the UMTS Terrestrial Radio Access Network (UTRAN). A RNS contains one Radio Network Controller (RNC), which is responsible for resources, and one or more node Bs, which are responsible for radio transmission/reception in one or more cells [21].

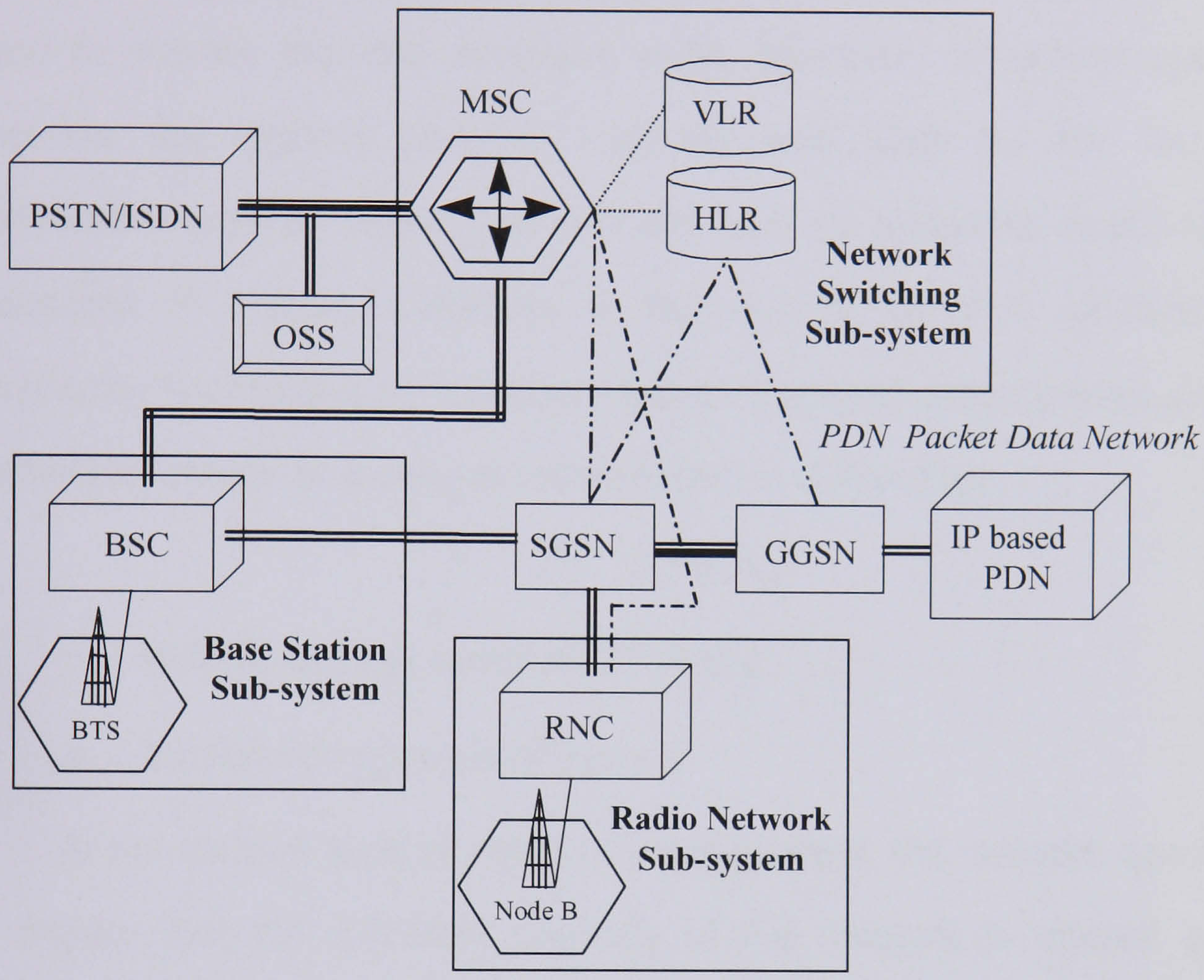

Figure 2-7 *Proposed structure of UMTS*

2.1.3 System Efficiency and Capacity Optimisation of Cellular Networks.

As mentioned above, cellular networks consist of two separate components: a mobile network and a fixed support network. The capacity of the fixed support network can be increased easily and without significant cost to the service operator, but the total capacity of the radio interface is limited by the radio frequencies allocated to cellular telephony [22]. In addition, although the radio interface capacity can be increased through cell splitting (see Figure 2-1), this is very expensive. Thus, the bottleneck in a cellular network is the air interface and the efficient utilisation of the available radio resources is, therefore, paramount. At the highest level of network management, operators need to ensure that the available radio spectrum is utilised optimally and provides the highest possible capacity and QoS for the network. The benchmark against which the efficient use of available radio spectrum is measured in cellular networks is the overall spectral efficiency η. This parameter is obtained by considering both the modulation and multiple access method efficiency of a cellular network and is defined as:

$$\eta = \eta_a \eta_m \qquad (2\text{-}1)$$

η_a - Multiple access spectral efficiency;

η_m - Modulation spectral efficiency.

At the second level of network management, the network operators have to ensure that the available capacity of the network is utilised optimally in conditions of vastly fluctuating user demand (both spatially and temporally). In the following sections, the factors affecting the overall spectral efficiency and limitations to the capacity of the networks discussed above will be identified, offering a brief overview of the highest level of network management and planning. Then the teletraffic management problem of matching the available network capacity to user demand and optimising the utilisation of the network will be discussed.

2.1.3.1 Capacity and overall spectral efficiency of FDMA and TDMA systems.

Multiple access spectral efficiency is linked to the access technique employed by the cellular network to ensure that multiple users can access the network resources. Three multiple access techniques are currently in use: FDMA, TDMA and CDMA.

Analogue systems are accessed using a Frequency Division Multiple Access (FDMA) protocol. Each user is allocated a frequency channel (a fraction of the total bandwidth available (see Figure 2-8 (a)), which can be used exclusively for the entire duration of the call. For example, for AMPS by specification this is 30kHz per user, allowing a data rate per channel of 10kbps, whereas for ETACS the channel bandwidth is 25kHz, allowing a data rate of 8kbps. See Appendix A for further details on AMPS and comparison with other cellular systems.

Two multiple access techniques are used by second generation cellular networks: TDMA and CDMA. First, we shall look at the TDMA protocol as it can be compared directly with the FDMA technique explained above. CDMA will be explained later.

GSM and D-AMPS networks use the principle of Time Division Multiple Access (TDMA). Each user is allowed to use the whole bandwidth of a channel in the cell but only for a limited time slot. The slots of a number of users are organised into frames, which are transmitted over the air interface (see Figure 2-8 (b)). For example, in GSM the channel spacing is 200kHz (shared by 8 users) which allows a total data transmission rate for 8 users of 104kbps [12]. For D-AMPS, on the other hand, the carrier spacing specified by IS-54 is 30kHz, which allows a total transmission rate of 23.85kbps per three user signals[7] [12].

[7] The data transmission rate given excludes the signalling overhead also transmitted by the mobile stations in the normal course of their operation.

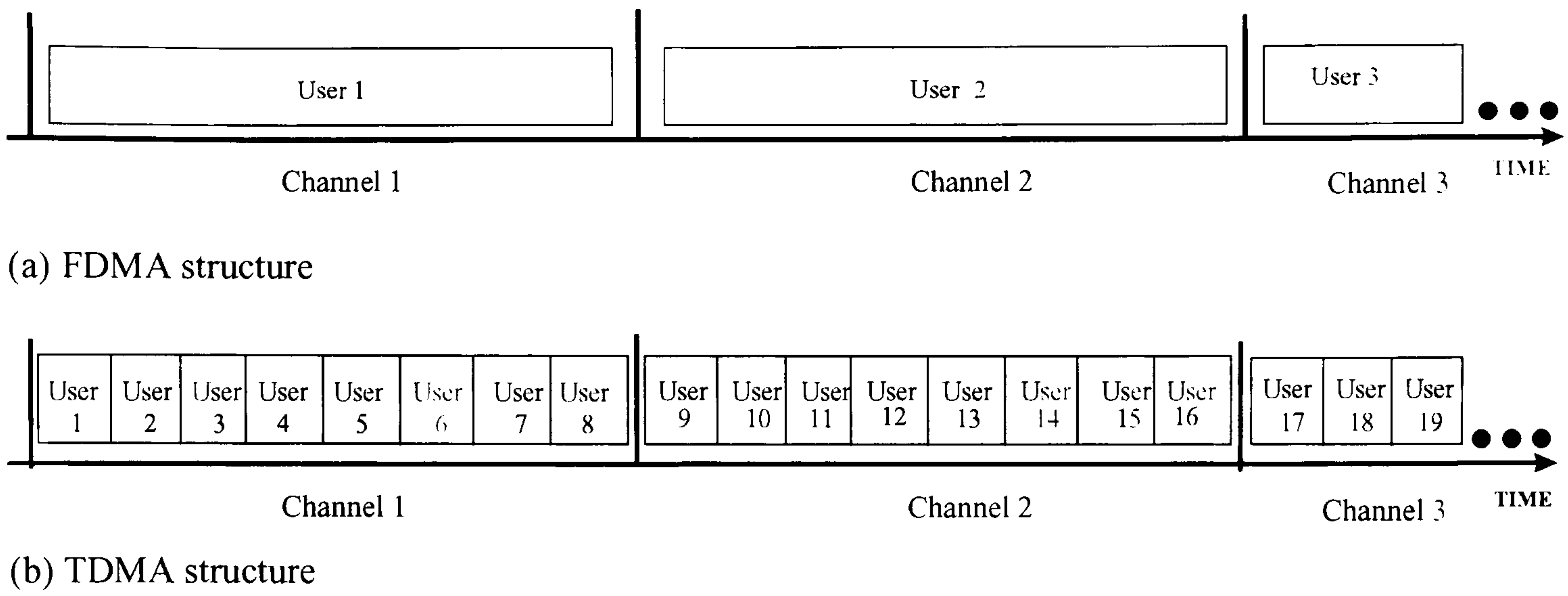

Figure 2-8 *A comparison between FDMA and TDMA.*

The multiple access effectiveness of a FDMA system η_a is defined as:

$$\eta_a = \frac{1}{K} \leq 1 \qquad (2\text{-}2)$$

K - Frequency re-use factor.

Therefore, η_a can be increased by reducing the frequency re-use factor. However, reducing the frequency re-use factor will lead to increased co-channel interference and reduced QoS.

By definition the total number of users supported by a FDMA or TDMA system is [23]:

$$N_u = \frac{B_w / B_c}{K} \qquad (2\text{-}3)$$

B_w - Total system bandwidth;

B_c - Bandwidth allocated per channel.

Substituting (2-3) into (2-2) defines the access efficiency of a FDMA system η_a in terms of supported users as:

$$\eta_a = \frac{B_c N_u}{B_w} \qquad (2\text{-}4)$$

This result suggests that by optimising the number of users in the network the efficiency of the access technique can be increased.

On the other hand, the multiple access efficiency of a TDMA system η_a is defined as:

$$\eta_a = \left(\frac{(\tau M_t)}{T_f}\right)\left(\frac{(B_c N_{u/c})}{B_w}\right) \qquad (2\text{-}5)$$

τ - Duration of a time slot;

T_f - Frame duration;

M_t - Number of time slots per frame;

$N_{u/c}$ - Number of users sharing the same time slot in the system but having access to different frequency sub-bands.

It can be seen that an increase in the multiple access effectiveness of a TDMA system requires fine tuning of mutually exclusive factors such as decreasing the frame duration without affecting the duration of individual time slots and the bandwidth allocated per channel.

The second component, which determines the overall spectral efficiency of a FDMA or a TDMA cellular network, is the modulation efficiency. This is defined as:

$$\eta_m = \frac{N_u}{B_w A_c} \qquad (2\text{-}6)$$

A_c - total coverage area.

By substituting (2-3) into (2-6) η_m becomes:

$$\eta_m = \frac{\frac{B_w}{B_c}}{\frac{K}{B_w A_c}} = \frac{1}{B_c K A_c} \qquad (2\text{-}7)$$

Therefore, the modulation spectral efficiency of a FDMA or a TDMA system does not depend on the total bandwidth of the system. It depends only on the bandwidth allocation per channel, the total coverage area and the frequency re-use factor[8].

2.1.3.2 Capacity and spectral efficiency of CDMA systems.

An alternative to TDMA, specified fully by the EIA/TIA IS-95, is Code Division Multiple Access (CDMA). In CDMA, all users in a particular cell use the entire frequency band allocated to this cell, rather than just part of it, as a channel [24]. Furthermore, CDMA uses a frequency re-use factor of 1 [25]. To distinguish between individual signals in a cell, CDMA uses 'spread spectrum' technology in which individual radio signals are coded using unique pseudo-random code sequences. Each user is assigned a "key" which is used to "spread" the signal across the entire band allocated to the individual user (say 1.25MHz as specified by IS-95). The receiver can only decode the signal sent by the mobile station with the same code as all other signals are dispersed as noise (see Figure 2-9) [26][9].

[8] The capacity of the network can be increased further by using Enhanced TDMA (E-TDMA) instead of the traditional TDMA. E-TDMA uses the natural pauses in every conversation for transmission of additional data, thereby increasing efficiency.

[9] This technique was first employed by the military to hinder signal jamming, and its advantage is that it offers a high level of security as intercepted calls are coded.

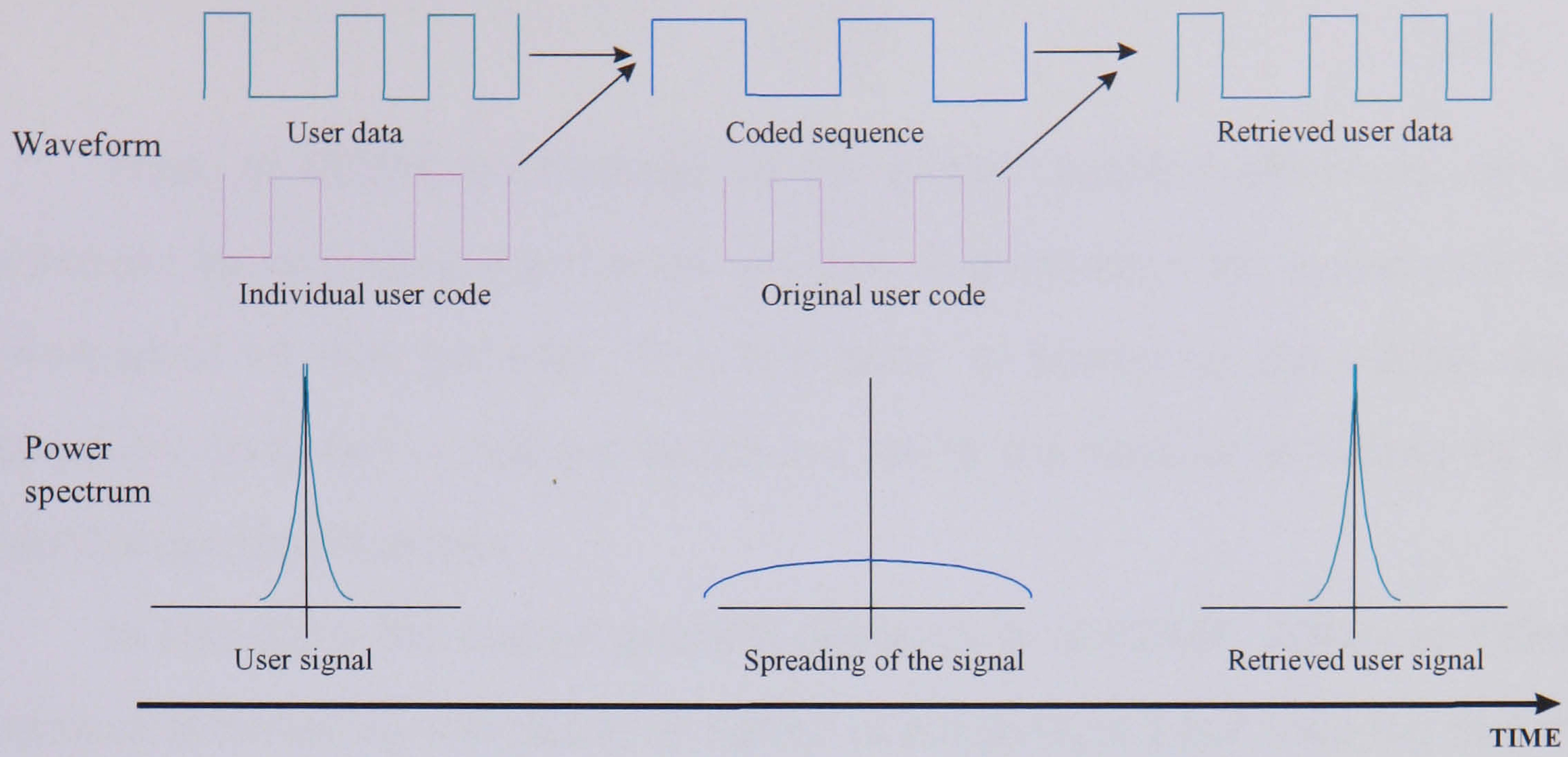

Figure 2-9 *Code division multiple access technique (CDMA).*

CDMA is a soft capacity system, *i.e.* there is no hard limit to the total number of users that the system will support. However, it is possible to estimate the number of users supported by a cell using:

$$N_u = \frac{\eta_f \eta_b c_d G}{v_f} \times \frac{B_w}{R \times (E_b / I_o)} \tag{2-8}$$

η_f - Frequency re-use efficiency;

η_b - Bandwidth efficiency factor;

c_d - Capacity degradation factor to account for imperfect Automatic Power Control (APC);

G - Number of sectors in site antenna;

v_f - Voice activity factor;

R - Information bit rate plus overhead;

E_b/I_o - Bit energy-to-interference ratio.

The overall spectral efficiency for a CDMA system (η) is defined as:

$$\eta = \frac{N_u R}{B_w} \qquad (2\text{-}9)$$

Thus, in CDMA an increase in the overall spectral efficiency can be achieved by increasing the number of users supported by the system and the information bit rate per user. The first factor is limited by the overall radio spectrum allocated to cellular telephony, while the second is limited by the technology specifications.

In summary, the overall spectral efficiency in a FDMA, TDMA or CDMA network is limited by the multi-cell nature of the networks but ensuring that the networks utilise the available spectrum optimally would allow network operators to support the maximum number of users. The fixed nature of the total available capacity imposes the question of the optimal utilisation of the available channels. This has to be done in conditions of fluctuating demand and user mobility. The next section will look at this problem and after introducing the necessary teletraffic engineering terms, will discuss the suggested techniques for optimal capacity utilisation for both 2nd and 3rd generation systems.

2.1.4 System Capacity Utilisation Measures.

The efficient utilisation of the available network capacity can be measured by either calculating the amount of traffic the network carries (using traffic flow or intensity) or by calculating the proportion of calls that were lost as a result of demand exceeding capacity (Grade of Service).

2.1.4.1 Traffic intensity

The traffic flow or traffic intensity in a cell is defined as the product of the number of calls during a specific time period and the average call duration (also known as call holding time)[10]. The period of time that is normally considered in traffic theory is an hour, so traffic intensity is the product of the call arrival rate λ (expressed in calls/hour) and the call holding time τ (expressed in hours/call). It is written:

$$E = \lambda\tau \text{ Erlang.} \qquad (2\text{-}10)$$

Measured in units of Erlang, traffic intensity specifies the load on the network as a whole or the load on the individual channels in a particular cell.

2.1.4.2 Grade of Service

Demand for cellular network services can fluctuate as a function of both the time of day and the user spatial distribution. An excess of user demand over available system capacity at any given location results in lost calls. Teletraffic engineering aims to deploy available network capacity to minimise the lost calls on a cell by cell basis. There are two terms for classifying those lost calls, depending on the mode of the mobile station when the call was lost. A mobile station in a cellular network can be in two distinct modes: dedicated mode, when a user is engaged in making a call; and idle mode, when the user has switched on their handset and is waiting for calls [17]. If the capacity of the network is saturated while the MS is in idle mode, the user will not be able to initiate or receive calls. This situation is described as *call blocking*. If a user moves to a fully saturated cell when the MS is in dedicated mode, there will be

[10] Call holding time is defined as the time for which a customer will use the service facilities. This is normally expressed as a probability distribution function with a mean and a variance. Recent research and empirical data courtesy of BT suggest that the call holding distribution time for cellular networks is log-normal, with mean and variance depending on the cell size, coverage area and the speed of the moving mobiles.

no free channel on which the call can continue. In this case the call will be *dropped*[11].

The total number of blocked and dropped calls is a major indicator of the performance of a cellular system and in teletraffic engineering it is defined as the Grade of Service (GoS) provided by the network. This is different from the concept of Quality of Service (QoS), which is a much broader concept and includes other parameters such as connection quality, voice clarity etc. The relationship between GoS and QoS is shown in Figure 2-10.

Cellular systems are designed with a GoS of 1% or better during peak-busy hour[12] [12], *i.e.* only one call in 100 should be dropped, so on average the GoS will improve when the system is not busy.

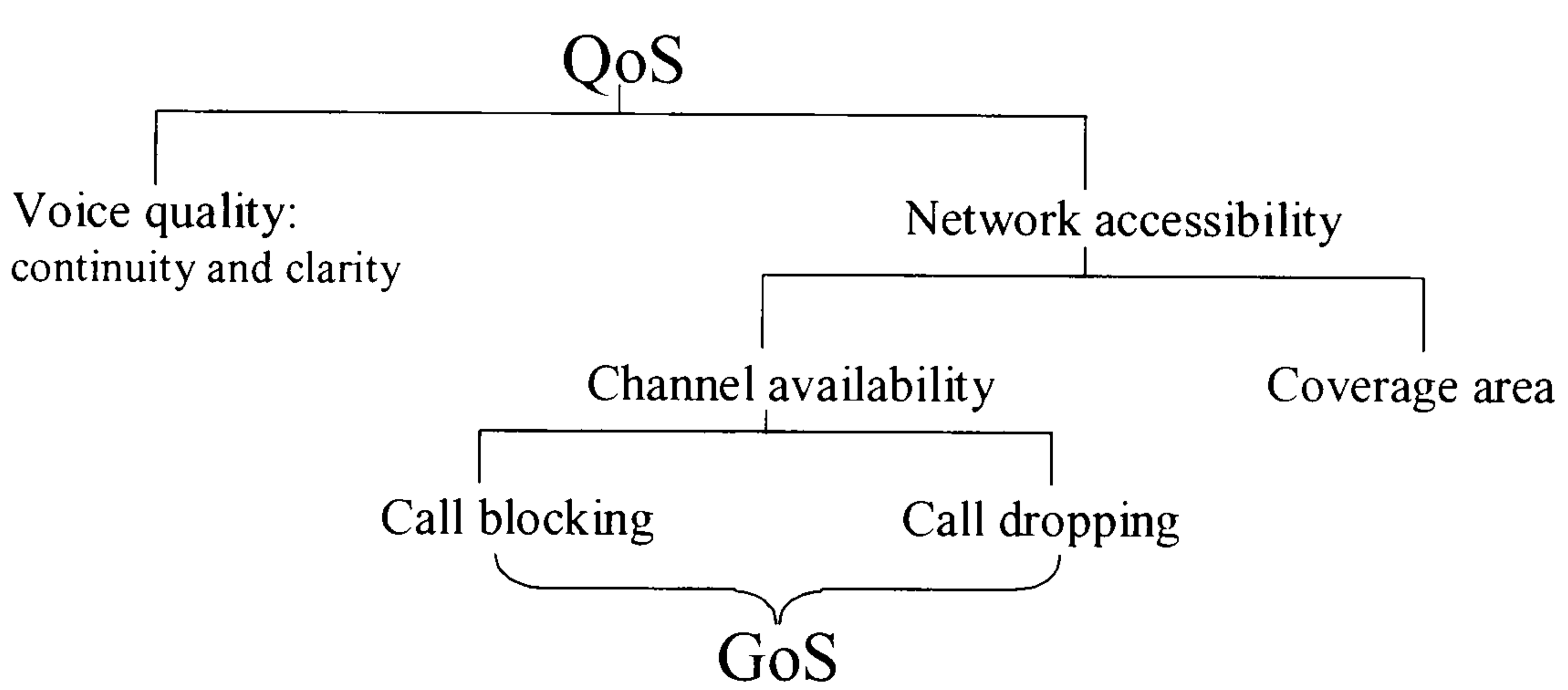

Figure 2-10 *Relationship between QoS and GoS.*

There are three formulae which enable teletraffic engineers to determine the number of channels necessary in a cell to provide given grade of service for a given traffic intensity: Erlang-B, Erlang-C and Poisson. The difference

[11] Research suggests that, overall, the probability of call blocking is higher than the probability of call dropping [27].

[12] Busy hour or peak-busy hour is defined as the period of 60 minutes during the day when the traffic intensity is at its greatest. This is averaged out over several weekdays and for cellular networks the peak usually occurs between 10 a.m. and 12 p.m. with a second peak between 1 p.m. and 3 p.m. The peak-busy hour changes at weekends.

between them is in the underlying assumptions as to the treatment of blocked calls. The formulae and the assumptions are compared in Table 2-1.

Formula	Blocking probability	Assumptions
Erlang B	$P_B = B(S,A) = \left(A^S\ S!\right) / \left(\sum_{K=0}^{\infty} A^K / K!\right)$	call arrival rate λ (Poisson) call holding time with mean $1/\mu$, traffic intensity $A = \lambda/\mu$ number of channels S blocked calls abandoned
Poisson	$P_B = P(S,A) = \left(\sum_{K=S}^{\infty} A^K / K!\right) e^{-A}$	call arrival rate λ (Poisson) call holding time with mean $1/\mu$. traffic intensity $A = \lambda/\mu$ number of channels S blocked calls held for the mean call holding time. then dropped
ErlangC	Probability of delay $C(S,A) = P_r[\tau_d > 0]$ with $C(S,A) = \dfrac{\left(A^S\ S!\right)\left[S\ (S-A)\right]}{\sum_{i=0}^{s-1} A_i / i! + \left\{\left(A^S\ S!\right)\left[S/(S-A)\right]\right\}}$ Probability of delay greater than t $P_r[\tau_d > t] = C(S,A) e^{-1(1-A)S\mu t}$ Average delay $E[\tau_d] = C(S,A)/(1-A)S\mu$	call arrival rate λ (Poisson) call holding time with mean $1/\mu$ traffic intensity $A = \lambda/\mu$ number of channels S blocked calls held until served

Table 2-1 *Comparison between the three traffic intensity formulae*

For the cellular networks operating at present, the first (Erlang-B) formula is perhaps the most suitable one because blocked calls are abandoned. In all three formulae, it is assumed that:

1. Calls occur at random intervals;
2. Number of users per channel is very large;
3. All users offer the same amount of traffic;
4. The system is in statistical equilibrium;
5. The traffic offered is known or accurately predictable.

As the Erlang formulae were developed to describe the system blocking probability for fixed networks, their extension to mobile networks is not always straightforward. For example, the velocities at which users travel will affect the performance of the cellular network and at first glance invalidate their applicability. If the velocity of the mobile nodes is increased while keeping all other factors in the system, *e.g.* cell size, constant, the *blocking* and *dropping* probability also decreases as shown in a study by Hong and Rappaport [27]. However, this result is consistent with the theory of the Erlang model because, due to the higher velocity, the overall call holding time in a cell will decrease, thus decreasing the overall system load. Although their underlying assumptions could be challenged, the Erlang formulas offer a useful tool for network management.

2.1.5 Traffic Profile.

Current user demand for cellular services exhibits some unfavourable properties: it is distributed very unevenly, both temporally and spatially, and poses problems to service operators attempting to optimise the use of available system capacity. There is a very significant difference in demand for cellular network resources during peak and off peak hours, as can be seen from Figure 2-11. A network geared to meet peak demand will be idle and have much spare capacity the rest of the time. In addition, traffic intensity varies spatially depending on the distribution of users in the network. The aggregation of a large number of users in a relatively small area (a football stadium or a movie premiere, for example) leads to the generation of particularly high traffic volume in that area. This phenomenon is known as a "hot spot" [28].

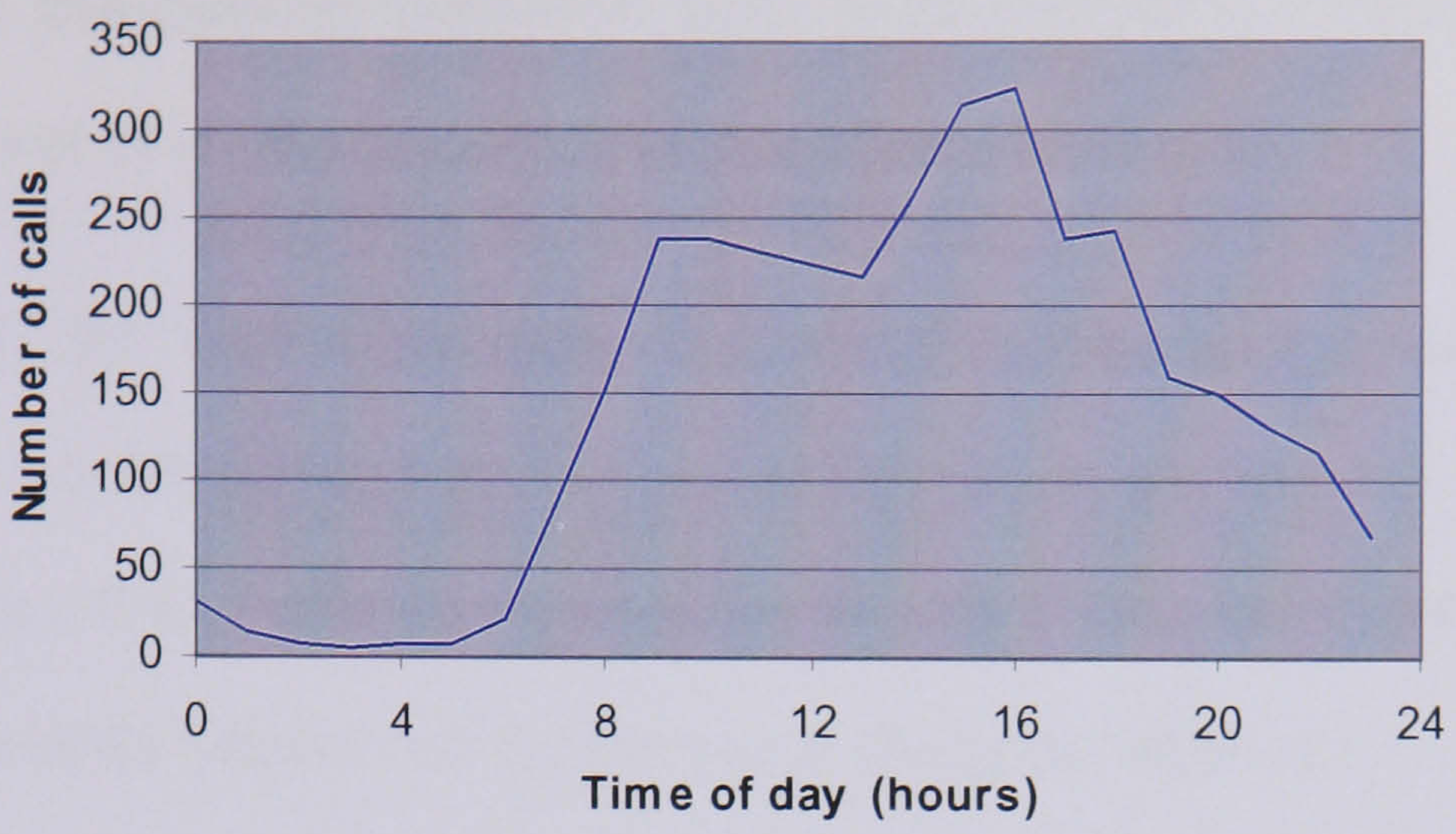

Figure 2-11 *Temporal distribution of cellular traffic in a small cell*[13]

2.1.6 Capacity Optimisation Techniques.

Total network capacity, as pointed out earlier, is limited by the frequency bandwidth allocated for cellular communications and certain network parameters chosen by the network operator (multiple access method, cell size, frequency re-use factor). Once the sizes of the cells are decided and the channels allocated the capacity on the individual cells is fixed. However, the fluctuation in demand makes some flexibility in cell capacity desirable. Two ways have been suggested for adding a degree of flexibility in the number of available channels per cell by either moving capacity between cells (in TDMA) or changing the size of the cells (UMTS).

2.1.6.1 Offering flexible capacity in first and second-generation systems.

There are two approaches that have been suggested for circuit-switched networks [29]:

- Moving capacity between cells (*dynamic channel assignment*);

[13] Data courtesy of BT.

- Using the overlap between cells and allocating calls to cells according to the level of traffic in each cell (*alternate routing*).

With the first approach, instead of permanently reserving channels for use within each coverage cell, the channels are kept in a centralised pool and are allocated to cells according to demand. The channels can be re-used in coverage cells separated by a re-use distance interval.

Channels could be allocated using a variety of different algorithms:

- First available: assigns the first available channel found in the search for free channels;
- Nearest neighbour: chooses a channel from the neighbouring cell.

Hybrid algorithms, which use some fixed and some dynamically allocated channels, can also be used. The negative side of *dynamic channel assignment* is the increased computational load on the system.

The second approach allocates calls to cells depending on their load, and various algorithms have been suggested:

- Last chance: a call blocked by the cell with the strongest signal is offered to the cell with the next strongest signal;
- Least busy cell: assigns overlap calls to the cell with the least number of calls.

Alternate routing strategies offer increased flexibility at the cost of some increased co-channel interference.

In conclusion, the available network capacity can be utilised more efficiently, and the number of blocked and dropped calls reduced by the application of *dynamic channel allocation* and *alternate routing algorithms*. However, this is done at the cost of an increased signalling overhead, although the increase is proportional to the complexity of the algorithm.

2.1.6.2 Offering flexible capacity in third generation systems.

CDMA systems have "soft capacity" and as a result will not exhibit call blocking when full capacity is reached; instead, the offered quality of service will slowly degrade, possibly encouraging customers to shorten their conversations according to Elliot and Dailey [30]. However, this shift in quality can be very frustrating to users using data services, for example, where decreased QoS has a profound effect. Spilling *et. al.* [28] suggest that an increase in demand can be met through adaptive cell sizing, *i.e.* reducing the radii of busy cells while neighbouring BSs increase their coverage area. This approach can work very well when dealing with traffic "hot spots" in the network. However, its effect is limited if all cells in an area become congested at the same time.

2.1.6.3 Other congestion control mechanisms in GSM.

Access to the GSM network is through the Random Access Channel (RACH) and is based on the slotted ALOHA method where, if a call request is not answered after a timeout, the sender assumes that there was a collision and the call request is repeated after a random time period. Currently, when the system detects congestion at the access point of the network (the radio interface), it has three methods to allow it to control throughput [17].

The first method for traffic control involves the alteration of the parameters of the RACH which determine the allowed number of repeat call attempts per MS and the time interval between attempts. The second method involves a rejection of a call request with a message forbidding the MS to access the network again for some specified length of time. The third method is the most effective option and relies on the concept of an access class. It controls congestion by barring whole classes of MSs to access the network (cell) at busy times (except for emergency calls). To achieve this, subscribers are split into 10 balanced sub-populations and the access class a user belongs to is stored in their SIM card and cannot be altered. Special categories of GSM users are defined as "Very Important GSM Subscribers" (Mouly and Pauted, 1992:371) [17], including security services, public utilities and emergency services, which will have a lower probability of call blocking. An ordinary user has no control over their blocking probability.

2.2 Alternative Approach to Capacity Optimisation – Dynamic Pricing.

The traditional engineering approach to capacity optimisation attempts to map user demand to the resources in the network by shifting and adjusting the number of traffic channels in a given region. However, wide fluctuations of user demand in time and space, coupled with the difficulties and cost associated with providing a flexible channel allocation scheme, lead to network congestion and a reduction in the GoS supported. This thesis suggests and examines in detail an alternative approach to optimising the use of the available network resources; instead of mapping the network resources to demand, user demand would be pushed towards the desired level that can be met by the available network resources. The ultimate lever that a network operator can use to affect user demand is the price of the calls and in order to even out demand, the price of calls has to change dynamically in time. It will decrease when demand is low to encourage more usage, and increase when demand is high to discourage new users and to possibly to encourage existing users to shorten their calls. Plotted in Figure 2-12 is the number of calls generated in a cell as a function of time[14] and a two-tariff (with peak and off peak) price charged by an UK network operator [31]. It can be seen that the price of calls is high even when demand is very low at the start of the day and low at the end of the evening when demand is still high.

The aim of a dynamically priced network would be to eliminate this anomaly and to distribute the traffic carried over the network more evenly, both temporally and spatially, resulting in a more efficient utilisation of network resources. It is intended as a "soft line" approach to capacity utilisation. For example, in current cellular networks only one call in a hundred should be lost at the busiest time of day.

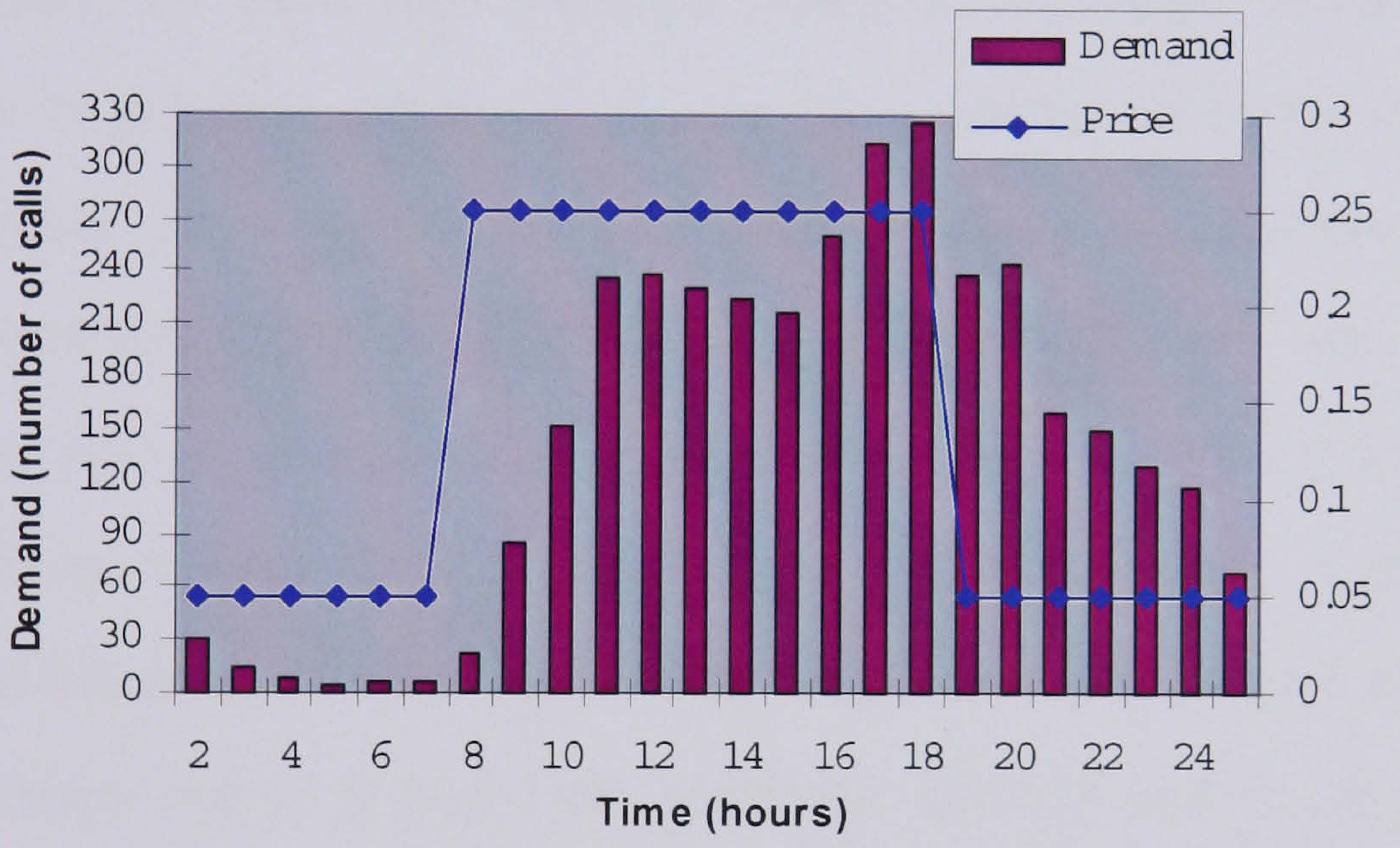

Figure 2-12 *Price fluctuations in a non-dynamically priced network*

In reality, however, this criterion would be difficult to meet due for technical reasons. If, for example, a network operator notices that a small cell in central London, with say 29 channels, has a blocking probability of 5% during the busy peak hour over a month, they might want to increase the number of channels in the cell. Using the Erlang B formulae from Table 2-1 to

[14] Data courtesy of BT

find the number of channels that will decrease the blocking probability to 1%, it can be seen that the number of channels needs to be increased to 32 [32]. In this particular example, only three additional channels would reduce the percentage of blocked calls to 1%. However, to achieve this by the "hard" approach of cell splitting, the network provider will have to incur significant capital costs in purchasing additional infrastructure and running costs for re-configuring the network.

2.2.1 Advantages of Dynamic Pricing Tariffs.

An advantage of dynamic pricing is that it will act as a natural control mechanism for calls with different priorities. As users have a choice as to whether to proceed with a call or not, the importance of the call will influence their choice. Thus, with dynamic pricing, a natural prioritisation of the calls will occur, ensuring that generally only low priority calls are lost. This will be an improvement on the current system, which has no means for regulating the importance of calls and so loses them indiscriminately. Furthermore, since users of the cellular network can be mobile and a user in a car, for example, would move out of a busy cell relatively quickly and could soon make the intended call at a lower price rather than postponing or not making it at all.

It is clear that dynamic pricing relies on the assumption that users will suppress their demand for the service when the network is busy and increase their usage at quieter times. Therefore, it is inevitable that a certain proportion of calls attempted during the busy hour and suppressed due to the very high price will never be made again. In effect, these calls are lost and represent lost revenue for the service provider. However, the proportion of calls lost due to suppression during busy periods has to be compared to the proportion of calls lost due to call blocking and call dropping as a result of system overload. From the point of view of the wealthier consumers the increase in price can be perceived as a special (premium) price to be paid for a better QoS. On the other hand, users who have to suppress or delay their calls as a result of the high price could perceive it a reduction in the QoS and therefore this issue has to be handled carefully by the network operator. In addition, the reduction in price when the network is underused would stimulate additional demand, which could compensate for the reduction in revenue due to calls lost due to the high price.

2.2.2 Dynamic Pricing Link with Control Theory and Economics.

The idea behind a *dynamically* priced network is that it will give feedback to its users based on the current state of the system, in order to steer the system into a desired state. The feedback sent to users will be in the form of different price for the calls for at that particular time in a particular cell. This approach shifts the problem from being a purely network engineering problem into the domain of control theory. Thus in order to analyse the problem and to find the optimal price function that will ensure that the system is taken to the desired state it is necessary to understand the basic principles of control theory.

In addition, changing the price of the calls will directly affect the revenue generated by the network operator. This is a very significant effect from a network operator's point of view and, therefore, also has to be examined.

2.3 Control Theory Background

Control theory deals with the analysis of dynamic systems and methods for developing controls to ensure that they remain in a desired state. In very general term the control problem can be defined as (see Figure 2-13):

"Given a system S, with measured signals y, determine a feasible control input u, so that a controlled variable z as closely as possible follows a reference signal (or set point) r, despite influence of disturbances w, measurement errors n and variations in the system" (Glad and Ljung (2000): pp 8) [33].

This problem is usually solved by letting the control input u be generated from y and r by a controller (or regulator) R.

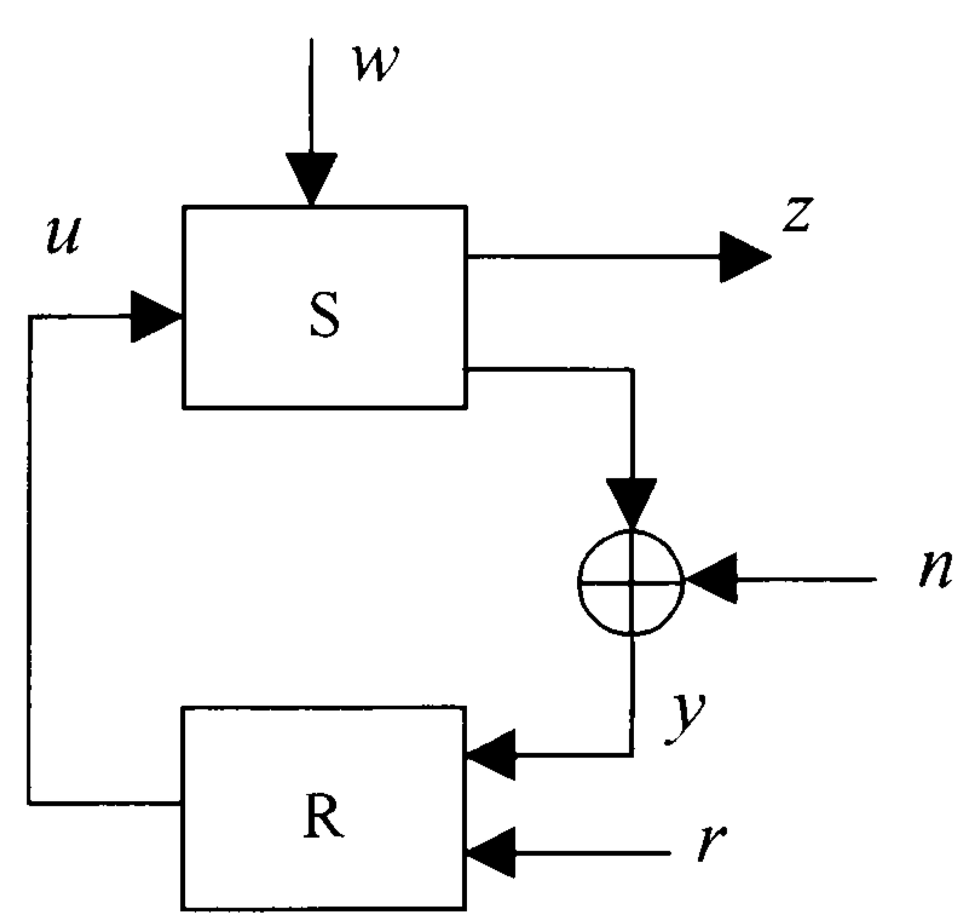

Figure 2-13 *Schemata of a controlled system*

Control systems can exhibit a number of properties and are usually classified depending on six of those. A system S is:

1. *Causal* if for every time point t_1, the output $y(t_1)$ only depends on the control $u(t)$, $-\infty < t < t_1$ and *non-causal* otherwise;

2. *Static* if for every time point t_1 , the output $y(t_1)$ only depends on the controlled input $u(t)$ for $t = t_1$ and *dynamic* otherwise;
3. *Time discrete* if the output y and the controlled input u are defined only for discrete time points[15] and *continuous* otherwise;
4. *SISO (single input, single output)* if for every time point t_1 the output and the controlled input are scalars and *multivariable* otherwise;
5. *Time invariant* if the mapping $y = S(u)$ is independent in absolute time and *time varying* otherwise;
6. *Linear* if S is a linear mapping, i.e.,

$$S(\alpha_1 u_1 + \alpha_2 u_2) = \alpha_1 S(u_1) + \alpha_2 S(u_2) \tag{2-11}$$

and *non-linear* otherwise.

The classification of a dynamically priced cellular system by those factors would depend on the assumptions made about the overall behaviour of the network users and the dynamic pricing function.

2.4 Chapter Summary.

After a historic overview of the development of cellular networks, this chapter looked at the structural and operational characteristics of current GSM systems as well as the future GPRS and IMT-2000 systems. The factors affecting the capacity and spectral efficiency of the systems mentioned above were identified, followed by definitions of terms commonly used in traffic theory for the measurement of network capacity utilisation. The temporal distribution of cellular traffic was presented, followed by techniques suggested for better utilisation of the available network resources in 2nd and 3rd generation

[15] Discrete control systems are described using difference equations, while continuous systems are described using differential equations.

systems, such as dynamic channel assignment, alternate routing and dynamic cell sizing, were identified and their drawbacks acknowledged. An alternative means for fitting demand to the available network capacity through dynamic pricing was then suggested, followed by a discussion of its advantages. Finally, a significant economic implication at the application of dynamic pricing in terms of its effect on network provider's revenue and the terminology of control theory briefly discussed.

Chapter 3

In the previous chapter, the problems encountered by limited network capacity and fluctuations in demand was identified. Dynamic pricing was suggested as a means for regulating demand and improving the optimisation of the available network resources. This chapter will focus on pricing from an economic point of view. The economic objectives of network operators will be identified, followed by a list of the tariffs currently used by network operators for meeting these objectives. A mathematical model developed to predict the effect of dynamic pricing on the network operators market share will then be presented. The choices faced by network operators in deciding on charging units for accounting purposes in static pricing strategies will be presented, before describing some practical implementations of dynamic pricing. Dynamic pricing algorithms will then be presented for voice and packet based networks before an in-depth discussion of issues arising from the implementation of dynamic pricing, such as the signalling overhead in the networks and the best position for the algorithm. Finally, possible complications due to the cellular networks' interconnectivity with the fixed network will be examined.

3.1 Economic Aspects of Network Management.

Once the network infrastructure is in place, a network operator's objective from an economic point of view becomes maximisation of the return on invested capital. This is achieved through increasing both the size of the subscriber database (market share) and the revenue per subscriber (RPS). The following sections will identify the factors affecting market share and then give an overview of the strategies employed by network operators for optimising the revenue per subscriber.

3.1.1 Factors Affecting Service Provider's Market Share.

The market share of each network operator will depend on historic factors specific to each country such as the number of competitors in the market and the underlying structure of the supply chain, factors that can be controlled by the network operators themselves and intervention of the regulator. For example, historically the UK market is very deregulated and competitive with four network operators and a number of service providers, which has encouraged competitive pricing policies[16].

A survey of mobile phone users in the UK, conducted as part of this research, indicated that the most significant factors affecting service providers market share are the price of acquiring a phone and the price of the calls. These were closely followed in importance by the network operator's coverage area[17]. The price of the calls was the deciding factor in the choice of 34.4% of respondents and therefore, it is reasonable to assume that, keeping all other factors equal, users would choose the service provider offering the lowest call charges. In practice this assumption will not be strictly correct, however, as users have different preferences and, for example, business users who are given mobile phones as part of their job are more concerned with the QoS and coverage provided by the network.

[16] See Appendix B for detailed information about the supply chain in the UK market and effect competition on price for the calls.
[17] See Appendix C for methodology and other findings.

Based on these findings the dynamic relationship between price of calls and the market share of a service provider has been plotted in Figure 3-1. As well as market share, in the short run, the price for services will affect the user behaviour and the demand for network resources. This will influence the revenue per subscriber recuperated by the network operator and in the long run will affect their profit and the amount of investment into new network infrastructure. In its turn, capital expenditure would again affect the price the network operator charges. The relationship between the price for services and the expected market share of the network operator is very complex and the importance of pricing strategies for the operators' successful market penetration and share retention cannot be underestimated. Customer churn would be a particular problem in mobile markets where the penalties for changing network providers are relatively low.

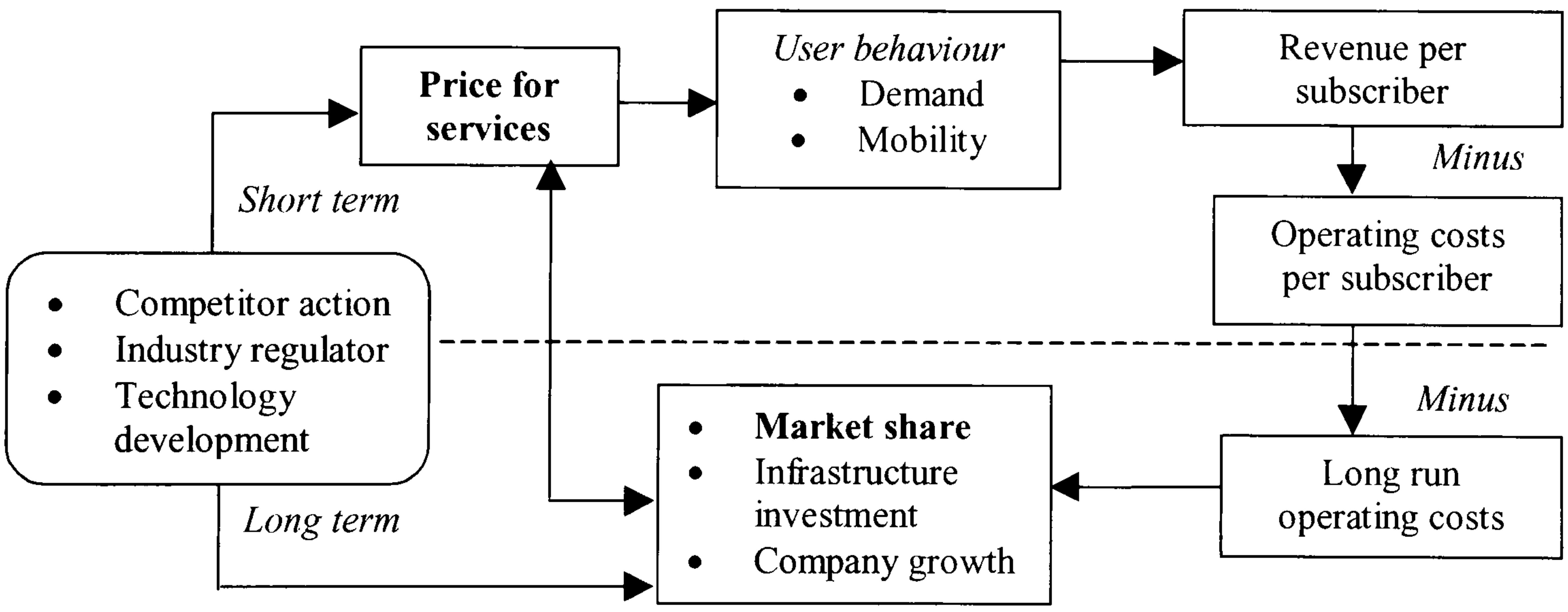

Figure 3-1 *Factors affecting network operator's market share*

For example, T-Mobile[18] attempted to enter the market in September 1993 by marketing itself as the operator which provided the highest quality of service on its network, initially limited coverage and simple and easy to

[18] New name for One 2 One since April 2002.

understand tariffs. In fact, to encourage demand, T-Mobile offered their customers all off-peak calls for free [11]. As a result, demand in off-peak hours increased dramatically and caused problems with system overload. This affected T-Mobile's reputation for offering good quality of service. Furthermore, under competitive pressure from other service providers, T-Mobile had to change and complicate their tariff structure in an attempt to attract more business users. Clearly, pricing is an important factor in determining the success of network operators in the competitive cellular telephone market.

3.1.1.1 Market segmentation.

As pointed out above, the gaining and retention of market share (or customer base) is the main goal of any service operator and the most effective means for acquiring new customers is through market segmentation. Market segmentation refers to the artificial subdivision of the market into different consumer groups, which allows the service provider to develop different marketing strategies to reach those groups. There are various bases for segmentation and the main criteria for effective segmentation have been identified by Kotler [34]. In general, each identified segment should be characterised by a set of common customer requirements, have measurable characteristics, common competitors and be of reasonable size (Ward [35]). Traditionally, telecommunication markets have been divided into business and private users. This particular segmentation is partly on the basis of the different price sensitivity of these two consumer groups. Alternative segmentation criteria could be:

- Geographic factors: depending of country size, city size and population density or separate units such as nations, states, regions, countries etc.

- Psychographic (or socio-economic) segmentation: this grouping is based on social class, life style or personality. For example, the lifestyle of a customer can influence their demand for telecommunication services.
- Behaviouristic segmentation: this segmentation divides customers into groups depending on purchasing behaviour, loyalty, usage rate and benefits sought.
- Demographic: consists of dividing the market into groups based on variables such as age, gender, family size, family life cycle income, occupation and education. Demographic criteria are among the most popular bases for segmenting user groups. For example, the need for communication services and usage often follows closely demographic variables such as age or occupation. In addition, these factors are easier to quantify than the psychographic or behaviouristic factors identified above.
- Size and profitability segmentation: a common phenomenon in many industries, including telecommunications, is that a small proportion of firm's customers account for a large proportion of their profit (also known as the Pareto effect) [35]. This is a useful basis for the segmentation of business customers, for example.
- Service segmentation: the users can be segmented on the basis of the services they require, for example, e-mail, web access, video conferencing etc[19].

In addition to the single criteria segmentation identified above, it is possible for the service operators to use a combined or tiered segmentation

[19] Users can also use service segmentation by differentiating between different service operators on the basis of the services provided. The UK Office of Telecommunications (OFTEL) has suggested that mobile cellular phones provide three types of service - access, incoming calls and outgoing calls. Most network operators do not consider access to the network, defined as "the capability to make and receive calls" (Cruickshank, 1998: 24) [54] as a service. In fact, only one UK operator offers a QoS guarantee, giving the users a free minute every time their call is dropped.

approach in which the users are divided depending on one criteria – for example, size or profit, and then subdivided again, for example, depending on their service requirements.

3.1.1.2 Tariffs for market share retention and revenue per subscriber generation.

3.1.1.2.1 Voice calls tariffs

Examples of pricing packages used by current network operators to increase their market share and the RPS include: bundled tariffs, single rate tariffs, home or friends and family discounts, off–peak tariffs and any combination of the above. With the increase in data transmission services, data only tariffs are also being introduced by mobile operators [36]. A separate pre-paid option was also introduced and the subscriber base using this option has grown significantly. On average, cellular European companies have 8 tariff packages (a maximum number of 22) while Asian companies have on average 13 pricing packages [37].

These pricing packages and their advantages to users and network operators are examined below.

Bundled tariffs include a certain amount of free off-peak talk time in the subscription fee. They are widely used as a means of generating a predictable amount of traffic per user. Over two-thirds of Western European operators use bundled tariffs and the number is increasing rapidly. The tariffs are well liked by users who can predict their bill fairly well. From the service operator's point of view, such tariffs are desirable because they are paid in advance and unused minutes are a significant source of profit. All four UK network operators offer bundled tariffs [38].

With ***single rate*** tariffs, subscribers pay a single flat rate for all types of call (to fixed networks or mobile telephones) even when roaming (not connected to their own mobile network). These tariffs are aimed at a small sector of highly mobile business users and are deployed mainly by USA operators. With their universal charge, these tariffs overcome call inhibition due to price uncertainty about roaming minutes.

Tariffs utilising the ***home zone*** concept offer highly discounted calls to a predetermined area (mobiles in the same cell, for example, or fixed network telephones in the same local area as the mobile at the time). The home zone tariffs are "origination-based" (based on the location of the user at the time the call is made) and actively encourage the generation of calls to the home zone. This allows mobile operators to compete effectively with the local loop area of fixed networks. Home zoning tariffs in the UK are offered by Cellnet and One2One.

In contrast, tariffs utilising the ***friends and family*** concept offer discounted calls to nominated telephone numbers and are "destination-based" (depending on the destination to which the call is made). As well as encouraging loyalty, these tariffs are very attractive to mass market users and stimulate calling activity. In the UK, only Cellnet offers this type of tariff.

Off–peak tariffs were introduced because, as explained earlier, networks are configured to meet peak-hour demand; and operators, in order to increase their revenue, want to stimulate off-peak usage. As a result, most tariffs offer discounted calls during the off-peak period. These tariffs are attractive to consumers because they address their fears of the high cost of using mobile telephones. From the network operator's point of view, they are desirable because they attract new traffic. Such traffic could be incremental (calls which otherwise would not have been made on the mobile) or time-shifted (calls which were timed to coincide with the lower tariff rate, and otherwise would have been made during peak hours).

The increase in the use of data services over mobile terminals, in particular in the younger generation (in some cases up to 90% of their communication is based on SMS messages (Short Messaging Service), for example [39]), has encouraged network operators to introduce special rates for data. There are even packages offering "**data only**" services and barring the user from making voice calls. These tariffs are relevant to specific groups of users who use portable computers or personal digital assistants. At present, they have little effect on overall traffic, but they will play a significant role when GPRS is introduced, as GPRS will offer higher data rates and a constant connection to the Internet. In Section 3.1.1.2.2 below, the tariffs currently used for pricing WAP and GPRS will be discussed in detail.

A new ***pre-pay*** option, also known as ***pay as you go***, was introduced recently. The users buy vouchers for the value they want and can charge their mobile phone bill. They are not offered any free minutes or reduced off-peak rates but are encouraged to use their phone by getting a discounted price for all phone use after, say, the first two minutes a day. Currently, pre-paid cellular mobile subscribers comprise 23% of all subscribers world-wide and GSM pre-paid subscribers are expected to rise to 53% by 2003 [40]. All four UK network operators offer this package. The option is very attractive to users who worry about running up high bills and the pre-pay option is also popular in other industries. Research into the holiday industry, for example, has shown that people prefer to pre-pay for their holidays, and then enjoy them without worrying about cost [41]. Another advantage of pre-paid tariffs is that they can be used by users without credit rating (teenagers or people without bank accounts). The disadvantage of pre-paid tariffs from consumers' point of view is that they are normally higher than standard tariffs [42]. From network operator's point of view, the advantages of pre-paid tariffs are that they reduce the cost of customer acquisition and the risk of bad debts. The disadvantage is that they could reduce airtime usage (due to higher prices) and customer churn (due to ease of changing network providers).

To summarise, the competitive nature of the mobile cellular markets world-wide encourages flexible pricing tariffs, which attempt to capture and retain subscribers through satisfying the different needs of diverse user groups.

3.1.1.2.2 WAP, i-mode and GPRS tariffs.

It has been suggested that, with the introduction of data services, the pricing strategies of service operators will change significantly because, with packet-switched networks operators can charge for the amount of information transmitted rather than the time spent on line. In reality, however, service providers have been rather cautious about introducing different charging strategies. In the UK, Cellnet offer their business WAP customers price per minute Internet access at 10 pence or 9 pence per minute depending on the calling plan [43].

The Japanese i-mode service operator NTT Do Co Mo charges according to the volume of data transmitted, not the time spent on-line [44], at a rate of ¥0.3 ($0.0024) per packet of 128 bytes [45]. In addition, a monthly subscription flat rate fee is paid by customers who wish to access certain i-mode Internet sites.

The Australian service operator Mobilkom launched, as it claims, the first commercial national GPRS network at the beginning of August 2000. The service costs 0.99 Australian dollars per minute at peak times and 0.49

Australian dollars per minute at off-peak times for business users, and 1.99 Australian dollars and 0.99 Australian dollars per minute respectively for all other users [46].

These examples indicate that service operators are very tentative about introducing new charging strategies to customers, even if the new strategy is clearly to the user's advantage, as is the case with data services which can charge for the amount of data transmitted rather than length of the connection.

3.1.1.3 Effect of pricing on network operator's market share

Tools used in game theory can be employed for predicting the implications of a pricing tariff on the market share of a network operator. It has been used to compare the effect of both fixed and dynamic pricing strategies. Fishburn *et al.* [47] compared the market shares of two companies, one of which charged a fixed fee, and the other one charged per-use fee and concluded that per-use pricing does not have a definite advantage over fixed fee pricing. Although they argue that in the long-run fixed fee pricing would be a winner, they also point out, that in conditions of competition, price wars are very likely, which will drive the prices of both firms down, diminishing the advantages of fixed price pricing. MacKie-Mason and Varian [48] estimated the cost of dynamic pricing for the fixed line Internet and showed that, in a competitive market environment, optimal dynamic pricing is economically more efficient, compared with a physical capacity expansion to satisfy the same levels of demand.

To estimate the effect of pricing on the market share of a cellular network operator, a mathematical model developed by Mason [49], will be adapted to compare the expected welfare of users of cellular networks with and without dynamic pricing. The original model measures the consumer utility gained from using a particular firm in a competitive market place. It will look at two horizontally differentiated firms, which means that some consumers will prefer one of the firms and others the other, even if both firms offer identical products at identical prices. This differentiation is realistic because if, for example, a company introduces dynamic pricing, some of its existing customers can be reluctant to change their service provider. The degree of consumer loyalty will be determined by the magnitude of the parameter $m \geq 0$.

In addition to existing users, some new users will be attracted from the competitor. Effective tools for acquiring new users are advertising, sales promotions and discounts and the amount of these the network provider can afford will be proportional to the revenue of the network operator. In the market share estimation model this will be represented by the term wR_i with $w \in (0,1)$ a random number representing the degree of success of the network operators marketing campaign and R_i representing the normalised total revenue of operator i.

The total welfare of the users will depend on the quality of service that the users experience and the overall network usage of the respective network operator. The quality of service users experience will be calculated as the users' Value for Money (VFM) term defined as:

$$VFM = \frac{1 - P_{B_i}}{P_i} \qquad (3\text{-}1)$$

P_i - Weighted average price[20] of firm i;

P_{Bi} - Probability of call blocking of firm i.

[20] The weighted average price will be calculated by taking the probability of any one price value occurring, multiplying it by the price value itself and adding all weighted price values together.

As the probability of call blocking decreases, the VFM component increases (for constant average prices). The total amount of product consumed by the users in this case is the total number of calls completed in the network and these will be normalised and represented as N_i.

Assuming that the users are uniformly distributed along the interval $(0,1)$ and the two firms are located at the end of the line at 0 and 1 respectively [49], the total user utility from firm i for a consumer located at $0 \le x \le 1$ will be:

$$U(x,i) = m[ix + (1-i)(1-x)] + N_i + \overbrace{\frac{1-P_{B_i}}{P_i}}^{\text{User welfare}} + wR_i \qquad (3\text{-}2)$$

m - Location utility component;

N_i - Normalised total number of calls serviced by firm i;

P_{B_i} - Probability of call blocking with firm i;

P_i - Average weighted price of firm i;

w - Random promotion effect factor;

R_i - Normalised total revenue of firm i.

This expression will enable us to calculate the utility each user can gain from joining one of the firms and, therefore, will allow us to evaluate the effect of pricing on network operators market share[21]. The calculations will be done at the end of Chapter 6, after three dynamic pricing strategies are suggested and evaluated.

[21] It will be assumed that users will incur no additional costs when changing network operators.

3.2 Pricing and Charging: Definitions, Requirements and Background.

In the previous section the economic objectives of network operators were identified and the means available for increasing the consumer database and maximising the revenue per subscriber were described. In this section, the challenges faced by network operators in setting pricing tariffs will be discussed. The process of price setting consists of three stages: at stage one, the network operators have to choose the charging unit for quantifying and accounting for network usage. The second stage consists of a choice of type of pricing strategy, for example static or dynamic, linear or non-linear. The final stage involves the setting of appropriate pricing units to the chosen pricing strategy, for achieving the network operator's objectives.

3.2.1 Definitions.

Following the terminology of Kaussar *et. al* [50] the term *pricing* will refer to the process of setting a price on a service, on a product or on content, while the term *charging* will refer to the function that translates the technical values of used resources into monetary units. The term *accounting*, on the other hand, refers to the process of metering the use of resources by individual users.

Linear tariffs in telecommunications are used to describe pricing in which the user pays the same amount for all units purchased, *i.e.* a user will not receive a discount for buying more of the same service, although a company could provide many different services at different prices.

Non-linear tariffs can charge a user different prices for different quantities of the same service. These tariffs include two-part tariffs such as the time of day tariff in which users are charged different prices for peak and off-peak calls [51].

3.2.2 Requirements on Pricing Tariffs.

The discussion so far has looked at the economic objectives of network operators and showed the role of pricing in determining both operators market share and the RPS. The complex relationship between pricing and other economic parameters in the network mean that setting of the pricing tariffs needs to be done with considerable care. A 'good' pricing strategy has to meet the following requirements from the network manager's point of view (Stiller [52], or Ferrari and Delgrossi [53]):

(a) Encourage optimal network efficiency to gain maximum welfare and QoS. The pricing strategy has to encourage users to request network services and ensure network resource optimisation without blocking too many users due to over-demand;

(b) Achieve high probability of cost recovery. The pricing strategy should ensure that the network operators recover their investment costs. This requirement guards against the possibility that if prices are set too low the network operator may not be able to produce enough revenue to off-set costs and obtain a fair profit margin, even when the network is well loaded and efficiently run;

(c) Competitiveness of prices. The charges offered should be competitive with those of networks offering similar services;

(d) Low implementation and usage costs of the charging policy, *i.e.* the tools that need to be acquired for accounting and billing of customers and the overheads generated by such tools should be low.

From the point of view of the customers, the main properties of a 'good' pricing policy are:

(e) Comprehensibility: pricing policies should be easy to understand;

(f) Controllability: users should be able to control the total charges for communication by varying the length or QoS requirements of the connection;

(g) Predictability and stability: users prefer charges that they can predict, *i.e.,* they prefer to know what the total charge for a transaction will be beforehand and they also prefer prices not to change too often in time;

(h) Fairness: the pricing strategy should be perceived by users to be fair and reasonable.

3.2.3 Additional Factors Affecting Operators Pricing Policies.

As shown above, cellular network operators have to operate in a competitive market environment and as a result their pricing policies are further affected by government regulatory policies and other market forces. The main external factors affecting pricing policies in the UK are:

(a) Industry regulator OFTEL (Office of Telecommunications). The influence of OFTEL on the prices for mobile phones can be significant. For example, by changing the basis on which interconnectivity call charges were calculated in 1997 [54], OFTEL affected call prices from the fixed network to mobile phones[22].

[22] Interconnectivity charges refer to charges levied by cellular operators to fixed line operators for calls initiated in the fixed network and terminating on the cellular network. These charges are then passed on to the consumers. In the past, these were derived on a fully allocated historic cost basis (this type of accounting takes into account the cost of assets at the time that they were purchased). However, following a modification in the licences of the mobile operators, the charges of calls to mobile phones currently have to be based on long run incremental costs. This type of accounting, called forward accounting, takes into account the cost of the assets at a current time, together with the cost increment to deal with mobile calls. This has led to a reduction in the prices as mass production and new

(b) Necessity for future network investment. The price of mobile calls will be affected by the cost of 3rd generation licences and the need for installation of new infrastructure for the 3rd generation networks. It is expected that the significant investment that has to be made by the service operators will ultimately (especially with the rules imposed by OFTEL prohibiting cross-subsidisation of services) be reflected in the prices of handsets and calls[23].

As these examples indicate service operators may not always be able to implement the most efficient pricing policies for their network.

3.2.4 Charging Units in Telecommunication Networks.

The charging units used by the network operators to determine the charge a user has to pay can be based on different parameters of network utilisation. They include:

1. Pricing based on the usage of a unit of a particular network resource such as bandwidth, buffer space or amount of interference introduced by the user, for example (connectionless networks), or call duration (connection-oriented networks). This type of pricing can also be done on the basis of larger units such as per flow for data networks or per call (voice networks). It is known as usage based pricing;
2. Pricing based on access with users paying only a network access fee which includes all calls (flat rate pricing);

---

technologies have reduced the overall current cost of base stations and other equipment. As interconnecting charges make up the largest proportion (16.7%) of the accounted operating costs of network operators [37], the reduction of interconnect charges has a significant effect on call charges. In a similar manner the charges charged by the fixed line operators for terminating mobile originating calls will be reduced. Therefore, OFTEL plays an important role in determining the future prices of calls to and from mobile phones.

[23] Confirmed by Pat Gallagher, director of British Telecom Plc., who stated in August 2000: "We have to make up for the high cost somewhere. In the end we are not a welfare state" [20].

3. Pricing based on the level of congestion in the network (congestion pricing);
4. Pricing based on the time of day (time of day pricing);
5. Pricing based on destination of call (location oriented pricing for connection-oriented networks) or distance or number of hops to be travelled by packets (connectionless networks);
6. Pricing based on the Quality of Service or Type of Service requirements of users (Quality of Service pricing);

In determining the charging units on which the pricing will be based, the network operator may decide to use more than one of the options listed above. So, for example, a network operator may choose to charge every customer a flat access fee, and in addition to that, make individual users pay for their usage of network resources.

After deciding on the units that will underline the pricing strategy, network operators have to decide on the type of price or tariff that will be used. The pricing can be *static* or *dynamic*.

3.2.5 Static Pricing Strategies.

Static pricing strategies are characterised by the fact that the price for unit of resource remains constant even if the network performance does not meet the requirements of the network users or the industry regulator. Static pricing strategies for voice communication are currently used by network operators world-wide as outlined in section 3.1.1.2.1 above. With the development of data based services, however, the challenge has been to find the "best" pricing strategy for a packet based network. The static pricing strategies suggested in the literature for data and voice calls can be classified into five main types [55].

3.2.5.1 Flat rate pricing.

In this type of pricing discussed by Cocci *et al* [56], a price is set and applied on per byte of information transmitted basis. In essence, this is a usage-based scheme and users pay a sum to the network provider that depends on the number of bytes sent. This is a rather simplistic pricing model, as it does not account for flows in the network with different QoS requirements. If this model is applied to a GPRS network, for example, users with more relaxed delay requirements will pay the same as users with more stringent delay requirements and this will be perceived as unfair (see 3.2.2 (h) above).

3.2.5.2 Priority (differentiated) pricing.

Cocci *et al.* [56] extended their initial flat rate pricing model to include priority (*i.e.* service classes) for different services that will be carried over the network. Different service classes make different demands on network resources. For example, real–time services require stringent jitter and delay guarantees, while non-real time services are “best-effort”. It is suggested that it is fairer to charge higher fees for the bytes with higher priority, which are then assumed to receive better QoS. Priority pricing schemes have great potential as they ensure that the users pay proportionally to the network resources they require. Bohn *et al.* [57] further developed the class-based approach by suggesting static pricing schemes in which different classes in the network are guaranteed a *minimum* performance level.

3.2.5.3 Bandwidth based pricing.

In this type of pricing users pay for the amount of bandwidth they use and are charged more for higher bandwidth requirements. Clark [58] developed a model in which the users are expected to specify their requirements such as expected connection speed (bandwidth) and QoS, thus helping the network operator with the planning of the resource allocation.

3.2.5.4 Reservation based pricing.

Songhurst and Kelly [59] offered a charging scheme which takes into account not only the type of service and the amount of traffic the user has transmitted but also the length of the connection. The time element of the pricing scheme is introduced to represent the reservation component, *i.e.* the amount of time network resources are reserved for a particular network user rather than being available to other users.

3.2.5.5 Paris Metro Pricing.

This pricing scheme was suggested by Odlyzko [60] and is modelled on the Paris Metro pricing. In the past, Paris Metro operated in a simple fashion with 1st and 2nd class carriages that were identical in the number and quality of seats they offered. The only difference was the price of the tickets. People using 1st class paid more than people using 2nd class. As a result, 1st class carriages were less congested as only people who cared about getting a seat were prepared to pay more. This principle can be applied to telecommunication networks by offering users the same service at different prices. The main difference between this approach and the priority pricing schemes mentioned above is that, in this case, the users can choose which particular price to pay, whereas with priority pricing the network operator decides how the services should map onto the offered service classes.

3.2.5.6 Interference based pricing.

A possible unit for calculating the prices in a CDMA network could be the minimum interference that can be introduced by a user. As the requirements of different users can differ, the differentiation in the charges can be achieved through taking into account the bit rate requirements of the individual users and calculating the total interference or load they contribute to the network [61]. The load of user i is determined by:

$$\Gamma_i = \left(\frac{E_b}{N_o}\right)_i \frac{R_i}{W} \qquad (3\text{-}3)$$

R_i - Bit rate requirement of user i;

W - Total system bandwidth;

$\left(\frac{E_b}{N_o}\right)_i$ - Energy per bit to Noise ratio of user i.

This type of pricing will be perceived as fair by the users because it will ensure that they pay only for the resources they use. Different weighting of the nominal price can be introduced to represent the different QoS requirements for the different classes of service.

3.2.6 Dynamic Pricing Strategies.

Dynamic pricing strategies use price to manipulate user behaviour in order to optimise network utilisation, or any other parameter defined by the network operator. Compared to static pricing strategies, with a dynamically priced network the service operators have to decide not only on the charging units for accounting purposes, but also which parameter in their system to use as a control variable. Potential control variables for determining the changes in the price, depending on the type of network, could be:

1. System load (in terms of number of calls, or traffic intensity, packet load or total interference);
2. Percentage of blocked calls;
3. Packet delay;
4. Frame delay.

3.2.6.1 Practical application of dynamic pricing.

Dynamic pricing is not a novel idea and although it has not been applied to cellular telephony, dynamic (also referred to as real-time) pricing strategies have been utilised in the electricity industry. Study of their effect will enable us to make predictions about the expected effect of dynamic pricing in cellular networks. An alternative application of dynamic pricing over longer time-scale in the airline industry will also be reviewed and the conclusions summarised.

3.2.6.1.1 Electricity.

Real time (dynamic) pricing for electricity has been applied in practice for both industrial and residential customers in America and Europe and at least 13 US utilities and several European electricity suppliers have demonstrated that hourly pricing is feasible (Zarnikau *et. al.* [62]).

American commercial users were able to shift their usage pattern in response to real time rates and in particular during peak hours, although the response to real time pricing, was not uniform, with just two companies providing the bulk of the measured response [63]. Similarly, in the UK, the experience of Midland Electricity, reported by Dawn [64] showed that industrial consumers use the information from real time pricing to monitor their electricity costs more closely and to alter their working patterns. However, although 84% of Midland Electricity's industrial consumers used the information to monitor their consumption, only 53% used the information to manage their consumption [65].

Using real-time pricing on residential customers Electricité de France observed that the majority of participants paid less for their electricity as a result, although many did not look at the prices often enough to change their demand. The small sample of customers and the bias introduced by the optional nature of the tariff, however, do not allow generalisation of the results [66].

The practical implementation of dynamic pricing in electricity indicates that dynamic pricing is a feasible alternative to traditional pricing strategies; however the response of users to the dynamic prices is complex and variable with the majority of commercial users using dynamic pricing to monitor rather than manage consumption. Residential customers, on the other hand, could expect to pay less for their electricity when using dynamic pricing.

3.2.6.1.2 Airline industry.

The objective of airline carriers is to maximise the revenue generated from each plane, since once the plane leaves the ground all its free capacity is lost. Revenue maximisation subject to capacity and/or time constraints is also known as *yield management*. This is, in fact, very similar to the situation with cellular telephony or any other industry with limited capacity, as the resources are there and if not utilised, revenue is lost.

In order to maximise revenue, the flight operators change the price of the tickets dynamically depending on demand and the time left to departure. The operator offers available seats at the highest price they believe will be accepted by the customers at any given time. However, the dynamics of finding the revenue maximising seat allocation is complicated by the possibility of multi-leg journeys, over-bookings, discount allocation, passenger spillage, passenger upgrades from the discount class fare and demand interdependence (Belobaba [67], Brumelle *et al* [68] and Smith *et al* [69]). The application of DINAMO, a dynamic revenue optimisation program developed for American Airlines, shows that dynamic optimisation can significantly improve the revenue generated from the system [69]. American Airlines estimate their benefit from the dynamic pricing system at over $500 million per year.

Although the time scale of the optimisation problem in airline industry is different to the time scale of the optimisation problem in cellular networks, the problem of yield management still arises and can potentially be addressed by the application of dynamically priced tariffs.

3.3 Technical Implications from Dynamic Pricing Implementation.

3.3.1 Dynamic Pricing Algorithms.

3.3.1.1 Dynamic pricing algorithm for voice based networks.

A suitable control variable for implementation of dynamic pricing in voice based networks like GSM, is the degree of utilisation of the available capacity. If demand exceeds capacity this will lead to blocked calls and dissatisfied users. If, on the other hand, the available capacity is under-utilised this will result in inefficient use of the available resources and potential loss for the network operator. The utilisation of available resources accurately represents the state of the network, in terms of the available capacity and the QoS that the user receives from the network.

Figure 3-2 shows a flow chart of a suggested dynamic pricing algorithm developed for a GSM network. The algorithm is based on a regular update in prices as a function of the network load.

The update in price will affect deterministic demand $D(t)$, which is a function of the time of day and can be predicted fairly accurately, based on historic data. Another demand component is the random demand $S(t)$, caused by, for example, emergencies (Koschat *et. al.* [70]) but this type of demand will not be affected by changes in price.

The load or state of the network will change as calls start and terminate and a snapshot of the state of the network will be taken at pre-set time intervals to enable price updates. The price update interval does not have to be constant, although this will be the case considered in this thesis.

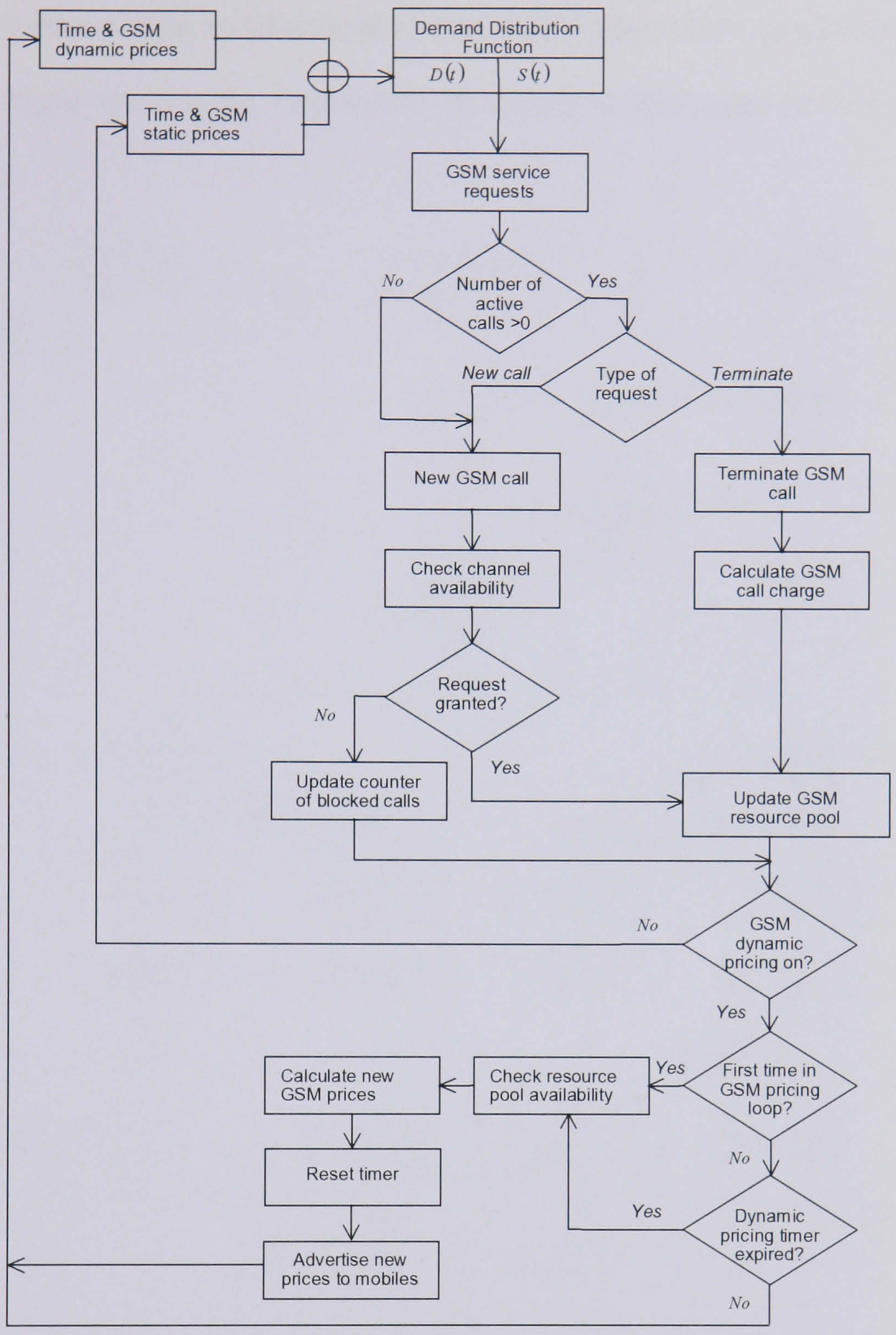

Figure 3-2 *GSM (voice calls) dynamic pricing algorithm*

3.3.1.1.1 Voice and data dynamic pricing algorithm.

GPRS networks are packet-based but they share the same resource pool with GSM networks and, therefore, a dynamic pricing algorithm has to take this into account. In Figure 3-3 a dynamic pricing algorithm for a GPRS-enabled network for both voice and data calls has been presented.

The left-hand side is identical to the GSM algorithm presented in Figure 3-2 but the right-hand side has been adapted to the requirements of packet based traffic.

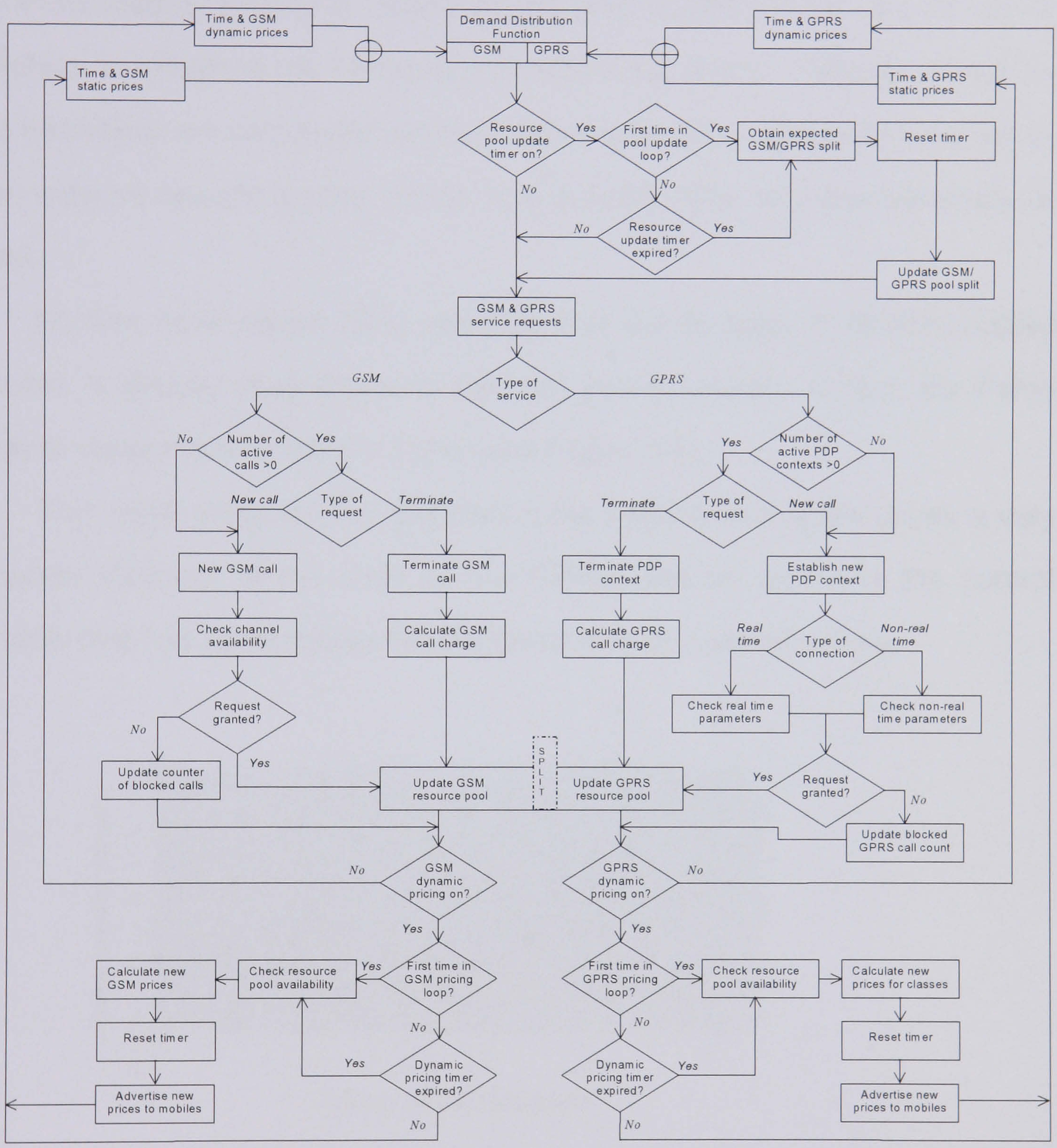

Figure 3-3 *GPRS (voice and data calls) dynamic pricing algorithm*

At first glance there is a significant difference between the control variables that can be used in GSM and GPRS. While in GSM based networks the percentage of blocked calls is the main criterion which has to be kept under control, GPRS-enabled networks have to meet the QoS requirements for reliability, delay, peak throughput, mean throughput and precedence as recommended by ETSI [71]. These are all good candidates for use as control variables to influence the changes in the dynamic prices. However, simplicity and tractability are also important requirements for price setting and, therefore, from network operators point of view it is desirable that only one parameter is used.

Studies have shown ([72] and [73]) that frame delay in GPRS-enabled network is directly proportional to the load in the network. In fact, the frame delay is unbounded for load > 5 kb/s (see Figure 3-4).

This result suggests that the load in the network can again act as a very accurate indicator of the state of the GPRS network and it is the control variable that has been chosen for the generation of dynamic prices.

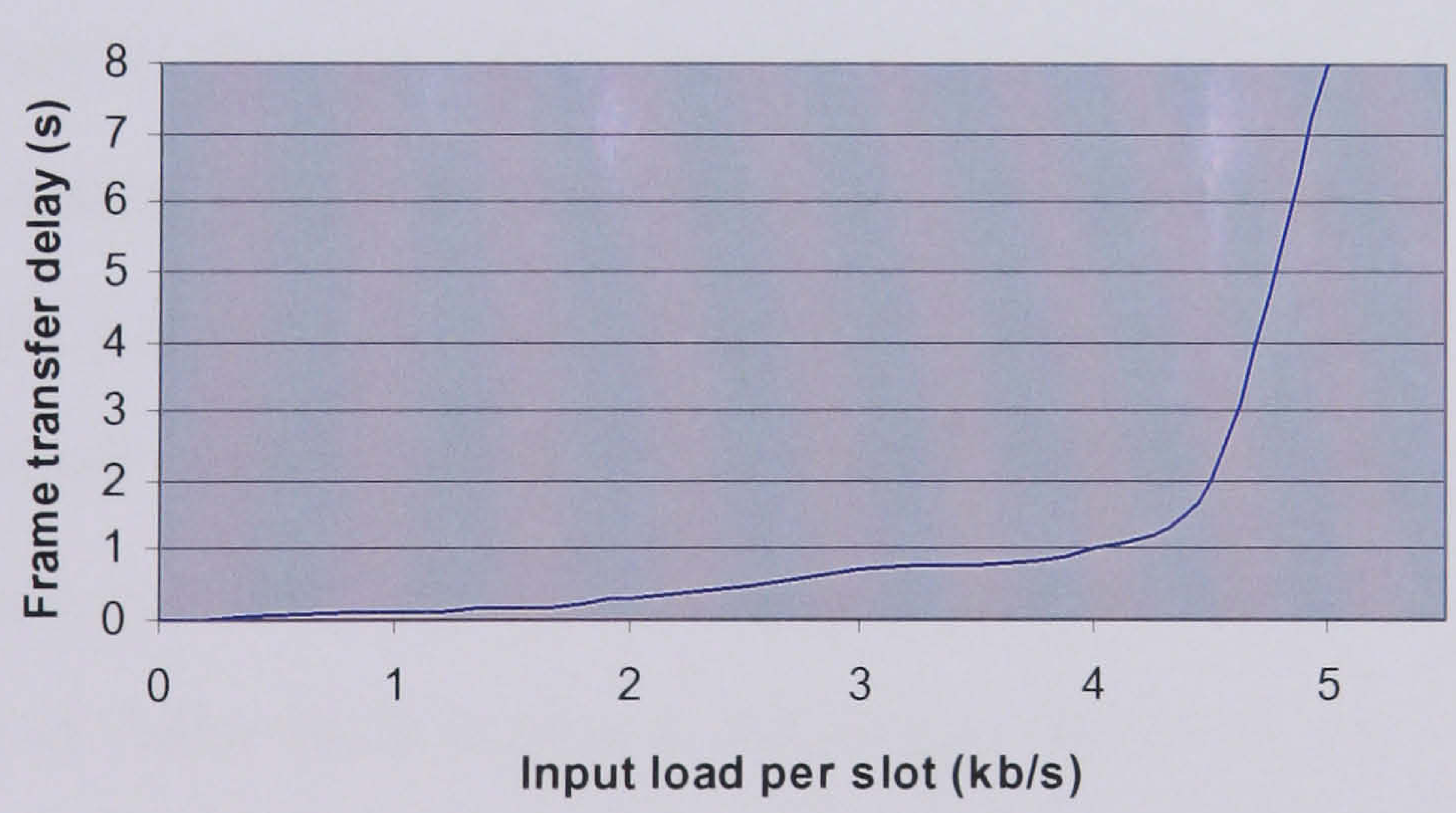

Figure 3-4 *Expected frame delay*

As users can request a variable amount of bandwidth, dynamic pricing for GPRS would be advertised on 'per time slot' basis, so if users want to increase their throughput, they can purchase additional slots. Furthermore, as real and non real-time applications have different service requirements, the network may not be able to meet the strict delay requirements for real-time traffic without "penalising" the non real-time traffic (Oliver and Songhurst [74]).

To reflect this, prices will be allowed to vary not only in time but also across the different service classes with different nominal (or per time-slot) prices being calculated for real-time and non real-time traffic)[24]. The term nominal price will be used to describe the charge per unit of resource at any given time.

The sharing of resources in GPRS between circuit switched and packet based services poses a very specific problem for dynamic pricing implementation. Industry regulations do not specify the exact nature of the division of resources and one possible implementation is to permanently subdivide the total capacity into two individual pools of resources. However, most users will have both GSM and GPRS enabled phones and will, therefore, be in a position to compare the prices of the respective pools. A large discrepancy in price between the two services can lead to users' choosing the cheaper option and as a result saturating the capacity and driving the price up. Meanwhile, capacity in the expensive alternative will be freed and the price will drop. This will now become the desirable alternative and users will shift to this type of service. The rapid and sudden shift in demand will lead to unstable system and lower QoS to users. As suggested in Figure 3-3, a solution to this problem can be the introduction of flexible divide between the two pools of resources. Thus the busier type of traffic is allocated more capacity thus

[24] For example, real time traffic will require a maximum mean transfer delay of less than 0.5 s per 128 octet Service Data Unit (SDU), while a SDU in the best effort category have unspecified delay. A SDU is defined in the reference model for OSI, as the amount of information whose identity is preserved when transferred between peer (N+1) -layer entities and which is not interpreted by the supporting (N)-layer entities [71].

keeping the free capacity in each pool proportionally the same. In this case, the nominal prices will be similar (if not the same) and the users will not have an incentive to shift their demand rapidly. For example, one nominal price will apply to all voice and real-time data connections and one reduced nominal price will be quoted for the non-real time data services. However, a drawback to this solution is that as an increase in intensity of one type of traffic, will lead to a reduction in the overall capacity of the network and thus indirectly increase the nominal price in the other type of traffic. This can be seen as cross service subsidy by the industry regulator and therefore this point has to be clarified.

The algorithm shown in Figure 3-3 can be applied directly to CDMA networks by taking the total interference in the network at any given time, instead of load as the control variable. This algorithm will allow for the co-existence of different types of access methods on one network, which is an additional advantage from the point of view of network operators, who may not want to discontinue existing services until the new ones have become established.

3.3.2 Dynamic Pricing Implementation.

3.3.2.1 Charging and billing overhead.

The limited availability of radio interface resources presents a problem for any algorithm requiring additional signalling or communication between the MSs and the fixed support network (BSTs, BSCs or the MSC). It could potentially use up valuable radio interface and reduce the throughput of the system.

The implementation of dynamic pricing algorithms relies on the dissemination of the new prices to the MSs. In GSM networks, the pricing information can be disseminated to the MSs in the relevant cell using the Broadcast Control Channel (BCCH)[25]. Only MSs in a particular cell listen to this channel and it has a free slot on the Downlink, which can be used to transmit the current call price to the MSs. As this implementation does not involve the use of additional resources, it will have no effect on the signalling overhead volume, *i.e.* the dynamic pricing would be transparent.

The effect of the implementation of dynamic pricing in GPRS network will differ only in the additional necessity to broadcast the updated prices to both GSM and GPRS-enabled mobile stations. In this case the broadcast to GPRS-enabled MSs can be done on the Packet Broadcast Control Channel (PBCCH).

The effect on signalling volumes between the fixed parts of the network from the implementation of dynamic pricing will be more significant. Presently, in GSM networks, the billing information is transmitted around the network by means of toll tickets [17]. Toll tickets are individual records created for each call that contain all the information necessary for calculating the call charges. They are created by the MSC/VLR and an average toll ticket contains information about the International Mobile Subscriber Identity (IMSI), the type of call (Mobile Originating, Mobile Terminating or Forwarded) and the call status (successful or not and if not, the reason for failure). The toll ticket also contains detailed data allowing the operator to accurately calculate the call charge (call starting and finishing time, number dialled, service type etc.). In a

[25] The Broadcast Control Channels (BCCH) is used in cellular networks for the continual broadcasts. on the downlink (base station to mobile stations) of information including base station identity, frequency allocations. and frequency hopping sequences.

dynamically priced network, in addition to all this information, it will be necessary to keep track of the price in the cell at the beginning of the call, and any subsequent price changes during the calls, which will make the toll ticket larger. However, the bandwidth available to the fixed support part of the network is not limited and can easily be increased, so the additional overhead generated by dynamic pricing is not significant. There will also be an increase in the processing overhead, although with the current speed of processors, this will not be significant.

These arguments apply equally well for a GPRS based network. Charging information is collected at both the SGSN or the GGSN and contains information regarding the volumes of data transmitted, the duration of usage of the packet data protocol addresses, the duration of PDP context, duration of usage of external data networks and location information. In particular the SGSN will collect information for each MS related to radio network usage, while the GGSN collects information relating to the external data network usage. This information will be transmitted from the relevant SGSN or GGSN to the network operators chosen Billing System (BSys) using the Charging Gateway Functionality (CGF) [75]. The Charging Gateway concept enables the operator to have just one logical interface between the CGF and the BSys. The information collected by the SGSN and the GGSN is collected in the form of Call Detail Record Charging ID (CDR). A CDR is generated every time a PDP context is activated and a unique Charging ID is assigned to all records generated for one PDP context. The addition of dynamic charging will cause an increase in the number of records generated for each PDP context. However, as in the case of the GSM networks the increase in traffic can be handled without significant decrease in the processing speed.

An additional problem with dynamically priced networks is the practicality of conveying the billing information to users in a meaningful way. Users require information to enable them to cross check their bills at the end of the month. Since the degrees of freedom in the network increase if prices and QoS requirements are allowed to vary, this information will increase significantly. A customer facing a comprehensive 30-page bill at the end of the month may not be able to interpret it accurately and this can lead to disputes and delays in bill payment. Therefore, network operators need to address this issue carefully. One solution could be to generate bills more frequently. The scale and expense of this task might be minimised by sending bills to the users via e-mail or text messages, for example, but this is unlikely to be without financial implications.

Another alternative could be the implementation of intelligent agents on users' terminals that collect and process information about the type of calls and the price of the calls users make and display it to users on request. The latter approach has the advantage of reducing the amount of signalling in the network and the processing overhead in the billing system. In addition, they may be very helpful to users by, for example, informing them when the price falls below a given threshold set by the user. This function could potentially lead to users swamping the network when prices suddenly become very low, following the pattern of the yo-yo effect described in section 3.3.1.1.1 above. However, in this case users will probably set different thresholds and dilute the effect. Further investigation would be necessary to determine the extent of this problem.

3.3.2.2 Positioning of the dynamic pricing algorithm.

Finally, the network operators have to decide on the best position for implementation of the dynamic pricing algorithm and the most suitable place in both GSM and GPRS networks will be the Base Station Controller (BSC). There are three main reasons for this choice. Firstly, the BSC is the common gateway to the radio interface for both the circuit-switched and packed-based services and, therefore, has full access to all data necessary for the evaluation of network performance such as system load, number of blocked calls, packet delay and revenue. In fact, some of this data is already collected by the network operators but is not used [17]. Secondly, there is enough processing power in the BSC to perform the calculations of the updated prices in finite time. And finally, the positioning of the algorithm in the BSC will minimise the increase in the overhead in control traffic in the network.

3.3.3 Implications of Cellular Network's Interconnectivity with the Fixed Network.

One reason why the application of dynamic pricing to cellular telephony could not be very straightforward is the wireless networks' interconnectivity with fixed networks. Calls terminating in a cellular network may have originated in the cellular network itself, in a cellular network run by a competitor, or a fixed network. Similarly, calls originating in the cellular network can terminate in any of the three types of network. In applying dynamic pricing, several issues arise. They include the issue of whether dynamic pricing should be applied to incoming or outgoing calls in a cell, or both. The capacity in a cell is limited and can be taken up by outgoing calls (the calls initiated in the cell) or incoming calls (the calls terminating in the cell). If dynamic pricing is applied to only one type of call, the effect of dynamic pricing will be lost on the rest of the calls. Three possible implementation scenarios are:

Case I: Dynamic pricing is applied only to outgoing calls. In this case, a proportion of all calls in the network will be unaffected by dynamic pricing. The overall effectiveness of dynamic pricing will be reduced and the degree of reduction will depend on the ratio of originating calls to the total number of calls.

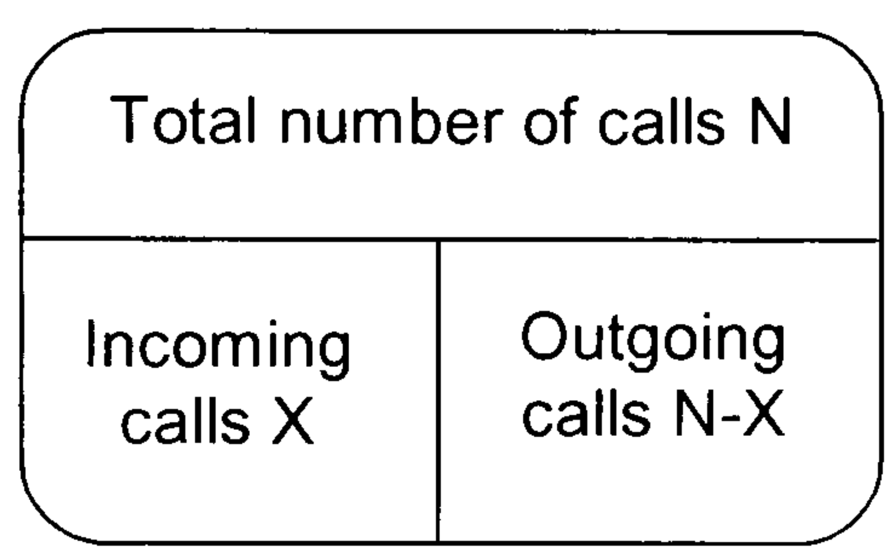

Figure 3-5 *Division of incoming and outgoing calls in a cell*

Case II: Dynamic pricing is applied only to incoming calls. This is a very similar situation to the one above but in this case the effectiveness of dynamic pricing will depend on the proportion of terminating calls to the total number of calls. However, two interesting issues arise. Firstly, in this case users can only see the price for the call they intend to make after they have dialled the party they want to call. This will inevitably generate some signalling overhead and if, after seeing the price, the user decides that they do not want to continue with the call, network resources will be used without any revenue being generated. Secondly, a percentage of the calls terminating in the cellular network originate in the fixed network, and communicating the current price to fixed line users can be difficult without a display on the terminal equipment. It has been suggested that 20%-30% of mobile terminating calls are made from the fixed network [76]. If these calls are also excluded from dynamic pricing and assuming a 50-50 split between incoming and outgoing calls in a cell, the effectiveness of dynamic pricing will be only 35%.

In both cases above the system is left open to abuse. If a party calls another party and the price they have to pay is quite high because they are in a busy cell or are calling a busy cell, they can always ask the called party to call them back. In this case, the proportion of incoming/outgoing calls will be increased and the effectiveness of dynamic pricing reduced. A solution to this problem would be to ensure that both parties see the same price. The implication of this strategy will be discussed below, for the case of mobile to mobile calls.

Case III: Dynamic pricing is applied to both incoming and outgoing calls. In this case, both incoming and outgoing calls will be subjected to dynamic pricing, but the system can still be abused if calls are made between two mobile phones and the price in one of the cells is lower that the other. Suppose that a party initiates a call at price P_1 in Cell 1 to a party in Cell 3 with price P_3. If $P_1 > P_3$, the calling party can ask the called party to ring them back, which will affect the proportion of outgoing to incoming calls.

Again a possible solution to this problem would be to ensure that both users see the same price. Implementation of this in practice will require the calling party to announce in advance their intention to call Cell 3, and to be given a price for the call after dialling the number.

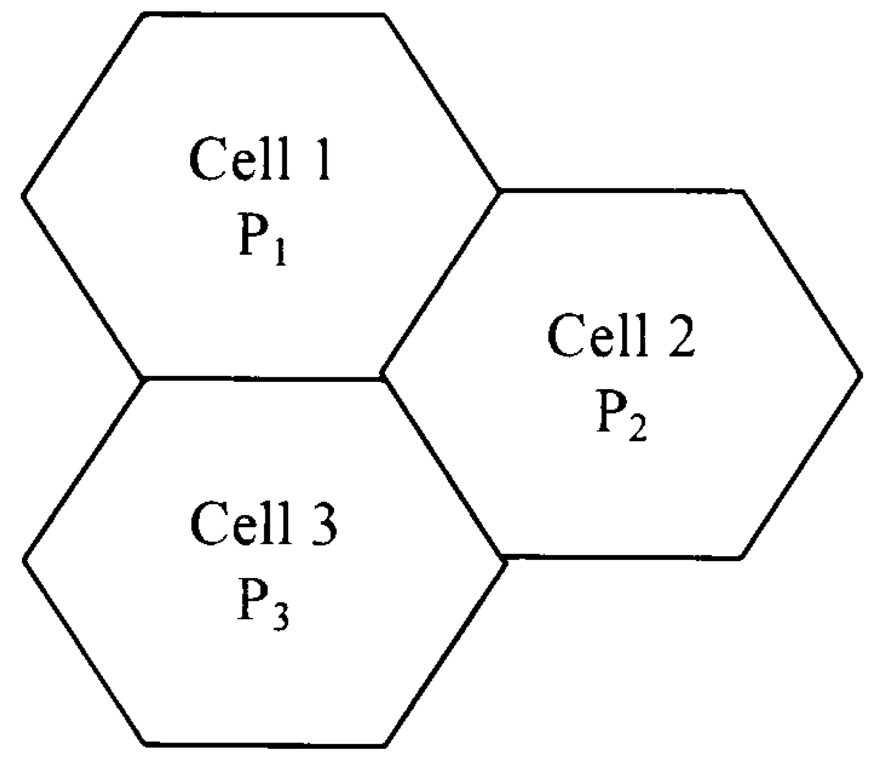

Figure 3-6 *Price distributions in cells*

The price advertised to the caller will be the higher of the two prices. This should be applied to all mobile-to-mobile calls to ensure that no discrepancy is created. It has to be noted that, if this strategy is adopted, a proportion of calls could be lost due to the fact that a higher than necessary price was advertised. This is very similar to the situation explained above in Case II.

The number of calls lost in this way can be estimated by using the tree in Figure 3-7. Assuming that 30% of all calls originate or terminate in the fixed network, and that the probability of going down any branch of the tree at each step is equally likely, around 12% of calls will be dropped on applying the higher of the two cell prices. This does not represent a significant proportion of the total number of calls carried over the network and the generated signalling overhead is also insignificant. In the above calculations, it was assumed that the users only ever see one price, the more expensive one. If users are able to see both prices at the same time, their own and the called party price, it will affect their choice. As a result, the probability along the tree branches will be different and affect the final result.

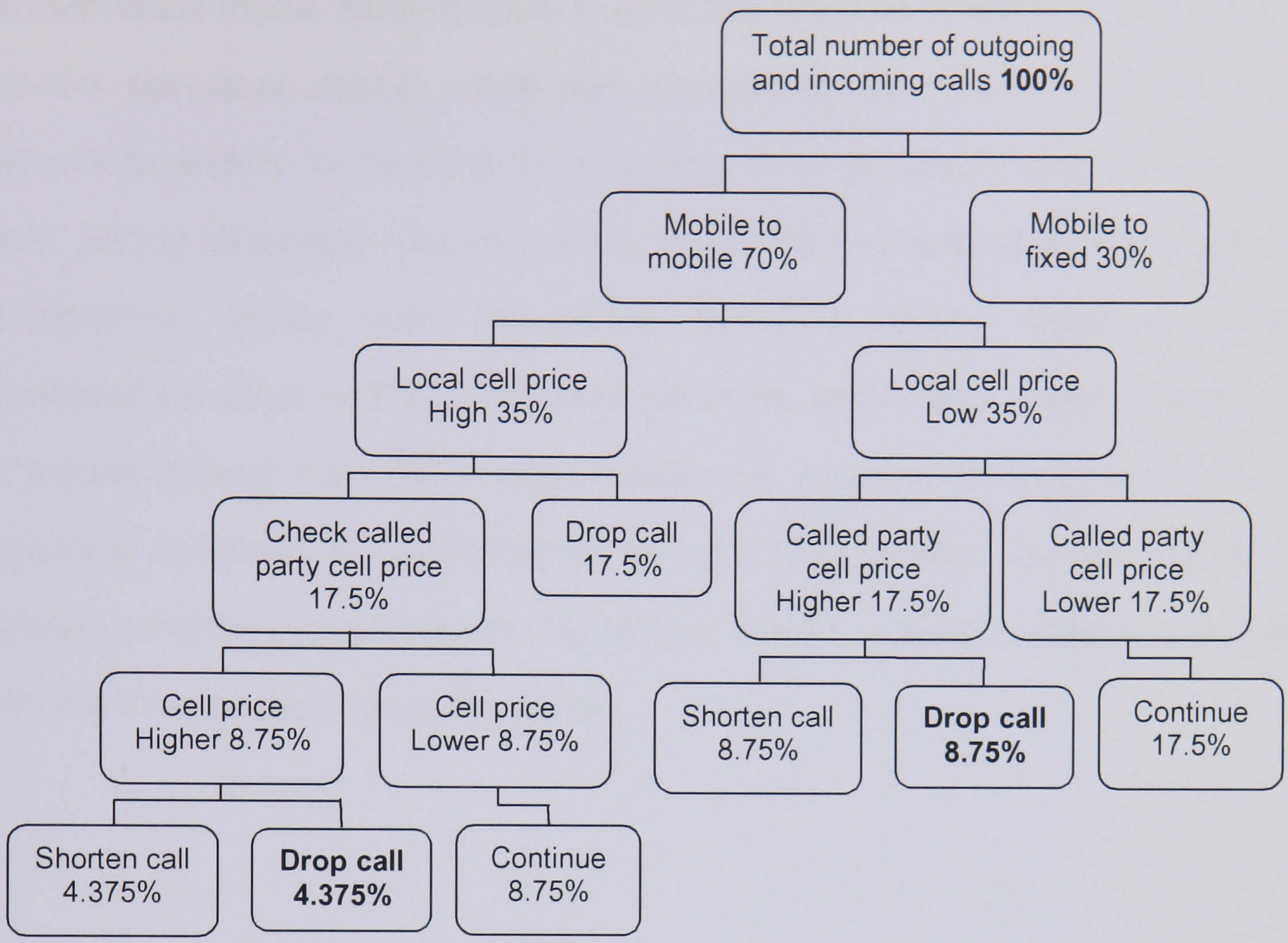

Figure 3-7 *Probability of lost calls due to applying the higher price*

Therefore, in order to be most effective, dynamic pricing has to be applied to both incoming and outgoing calls, and in the case of mobile-to-mobile calls the price of the more expensive cell has to apply. This will require the users to dial the party they want to call and then to look at the price and decide if they want to continue with the call. This may prove an inconvenience if a user is making a voice call, for example, but it is not going to be a serious problem if the user is making a data transmission call.

3.4 Chapter Summary.

This chapter focused on pricing from an economic point of view. The economic objectives of network operators were identified, followed by a list of the tariffs currently used by network operators for meeting these objectives. A mathematical model developed to predict the effect of dynamic pricing on the network operators market share was then presented. The choices faced by network operators in deciding on charging units for accounting purposes in static pricing strategies was presented, before some practical implementations of dynamic pricing were described. Dynamic pricing algorithms were presented for voice and packet based networks before an in-depth discussion of issues arising from the implementation of dynamic pricing, such as the signalling overhead in the networks and the best position for the algorithm. Finally, possible complications due to the cellular networks' interconnectivity with the fixed network were examined.

Chapter 4

The previous chapter looked at the economic objectives of network operators and discussed the implications of introduction of dynamic pricing from an engineering point of view. This chapter will begin with a discussion of the expected effect of dynamic pricing on the user behaviour. A comprehensive mathematical model for evaluating this effect will be introduced taking into account the price elasticity of user demand, as well as the existing pricing bias in the network and a substitution effect from the fixed network. In addition, a user mobility model showing the effect of price on user mobility will also be introduced. Then the discussion will focus on the final stage in any pricing policy: the setting of monetary units to the charging units. A literature overview of suggested optimal price setting strategies from an economic and consumer welfare point of view will be given and the drawbacks of the approaches highlighted.

4.1 Effect of Dynamic Pricing on User Behaviour.

The expected effect of dynamic pricing on the behaviour of the users in a cellular network will be threefold. Firstly, the changing prices will affect the number of calls the users make. Secondly, the prices will have an effect on the length of the calls the users make. So, a very expensive price could result in calls not being made at all, or calls being shortened. The third effect is specific to cellular networks and is directly related to the fact that users are mobile.

Faced with an expensive airtime, the users have a choice whether to proceed with the call or to delay it until they move to a less busy cell. Thus dynamic pricing will also affect the spatial distribution of users.

In the next section, the assumptions made about the behaviour of users will be listed and justified and a basic mathematical model to depict the effect of dynamic pricing introduced. The model will be developed further to consider some additional aspects of the system, such as the existing pricing policy in the network and user mobility as a function of price.

4.1.1 Assumptions.

Four assumptions will be made regarding the behaviour of users and the cellular system:

1. Users behave rationally, *i.e.* they consider the relative costs and benefits of any decision they make with their limited resources in hand. Therefore, as the price of a service increases they will purchase less of that particular service. This assumption is standard in classical economics theory (Cook and Farquhason [77]). It is also supported by results from a pilot user survey conducted as part of this project over the Internet (see Appendix C for survey methodology and other findings) and findings by Cosgrove and Linhart [78]. They studied the effect of pricing on demand for fixed telecommunication services and observed a positive correlation between increase in the price of calls and reduction in the total number of calls made by customers. This was true even for flat rate customers for whom reduction in call frequency had no effect on the bill. These findings confirm that the assumption that most users behave rationally is reasonable.

2. Price of calls is the most significant factor that affects user demand. This follows from assumption (1) and is confirmed by findings from the survey, which show that, the price of calls is the most significant factor in users' choice of service provider (see Appendix C).
3. We will consider a closed system: no mobiles will enter or leave the system. However, mobility within the system as result of pricing will be taken into account and this will be discussed in section 4.1.4.
4. We will only consider mobile originating calls to the fixed system and ignore all calls from the fixed network to mobile terminals and mobile-to-mobile calls. This will simplify the analysis of the system and the implementation of the dynamic pricing algorithm.

4.1.2 General User Demand Model.

The demand function of a product or a service shows the relationship between the price of that product or service and the quantity demanded by users at a given price. From assumptions 1 and 2 outlined in the previous section, it follows that the reaction of users to price will be rational, *i.e.* users will reduce their demand as price increases. Therefore, the demand functions that we will consider will be monotonically decreasing.

An important aspect of user demand is demand price elasticity. Demand elasticity is defined as the expected % change in quantity demanded divided by the % change in price (Varian [79]).

$$E_D = \frac{dQ}{dP_Q}\frac{P_Q}{Q} \tag{4-1}$$

Q - Quantity of product or service demanded;

P_Q - Price of the product or service.

By definition, if for a particular product $|E_D| < 1$ then its demand is inelastic, if $|E_D| > 1$ its demand is elastic and if $|E_D| = 1$ then the demand for the product is unit elastic. Demand elasticity depends on socio-economic factors such as income and age. In addition, the type of user business or private will also affect the magnitude of the demand elasticity. It is well established that business users have lower price elasticity than private users. In turn, private users can further be subdivided according to their disposable income into wealthy or poor, for example, which will again affect their respective elasticity of demand.

4.1.2.1 Types of demand as a function of price.

In economic theory, for the purposes of analysis simplification, demand functions are expressed as either linear (4-2) or exponential (4-3).

$$D_x = A + B_0 P_x \qquad (4\text{-}2)$$

A - Demand time constant;

B_0 - Slope of the demand function or demand elasticity;

P_X - Price of the product.

$$D_x = A P_x^{B_0} \qquad (4\text{-}3)$$

In practice, however, the shape of the demand function is likely to depend on the type of service. There is little published evidence on the shape of the demand function in cellular telecommunication networks and in theory any function that satisfies the assumptions outlined above could be accepted as the demand function. For example, the reciprocal function defined by equation (4-4) is potentially a candidate.

$$Q = \frac{A}{\sqrt[\beta]{P_Q}} \qquad (4\text{-}4)$$

Q - Demand for calls;

A - Demand time constant;

β - Elasticity coefficient factor;

P_Q - Price of calls.

A better choice, however, is a modified exponential function (4-5) suggested by Koschat *et al.* [80] because it was originally used for modelling demand in the fixed telecommunications network:

$$Q_t^k = q_t^k e^{-\beta_t^k p_t^k} \qquad (4\text{-}5)$$

Q_t^k - Demand for customers type k during hour t;

q_t^k - Random demand during hour t;

p_t^k - Price of a call during hour t for a type k customer.

This type of demand function will be modified to take into account behavioural features specific to the cellular network (the substitution and mobility effects are discussed in section 4.1.2.2.2 and 4.1.4 respectively) and so the simplified starting function for demand will be[26]:

$$Q = Ae^{-|\beta| P_Q} \qquad (4\text{-}6)$$

Q - Number of calls demanded;

A - Demand time constant;

β - Elasticity coefficient factor;

P_Q - Price of calls.

---

[26] To ensure that the demand function is monotone decreasing (assumption 1) we shall only consider $\beta \geq 0$.

This type of function has also been chosen because of potential ease of manipulation and properties that allow easy addition of the substitution and mobility effects.

The elasticity of this type of demand is (using equation (4-1)):

$$|E_d| = \beta P_Q \qquad (4\text{-}7)$$

So the elasticity of demand changes proportionally to both β and price as we move along the demand curve. Therefore, β can be interpreted as a calibration parameter, which determines the significance of the effect of price on user elasticity: a small β indicates relatively minor significance, while a large β means that the price will have a relatively large effect on elasticity. Thus β acts in a similar way to pure demand elasticity and can, therefore, be interpreted as a quasi-elasticity of demand.

The dependence of demand elasticity on parameter β poses the question of determining the appropriate elasticity for the demand. As the price in a dynamically priced network is not constant, it will not be possible to test the demand for any fixed value of the demand elasticity. However, by choosing the values of β while bearing in mind the interval of feasible dynamic prices, it will be possible to simulate a network with demand elastic or demand inelastic traffic. Although no information is available regarding the actual elasticity of demand for communication services, in this thesis it will be assumed that the demand is at most unit elastic[27] *i.e.* $|E_D| \le 1$. Further investigation of user behaviour is necessary for more accurate information on the elasticity of demand. The problem is compounded by the possibility of multiple services with potentially different demand patterns and elasticity. To

[27] Private communication with a network provider.

simplify the analysis, the demand model will be developed only for voice calls and the assumption will be made that the demand user elasticity β is uniform.

4.1.2.2 Other factors affecting demand.

The demand functions discussed so far give a relationship between the price and the expected number of calls made at that price. Plotting the expected number of calls at one provider's current peak (25 pence per minute) and off-peak (5 pence per minute) pricing rates, using the exponential demand function (4-6) and $A = 100$, the expected level of demand can be seen in Figure 4-1.

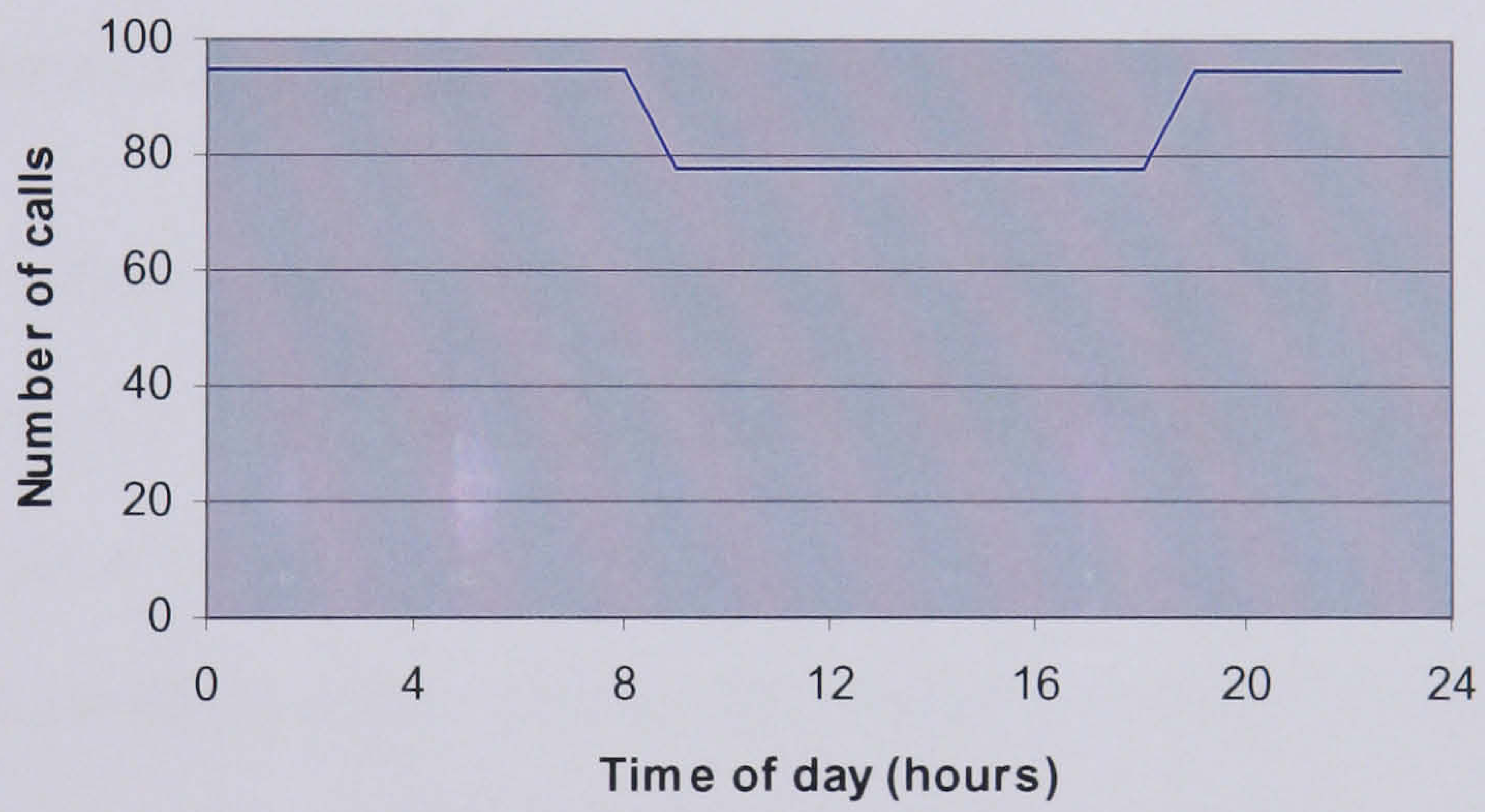

Figure 4-1 *Demand as a function of price*

Comparing this to the actual number of calls generated in a cell in real time in Figure 4-2 it can be seen that the demand pattern is very different. Therefore, price is not the only factor that affects the demand for calls. In order to make accurate predictions about the effect of dynamic pricing on expected user demand the model has to be modified further to match the actual data.

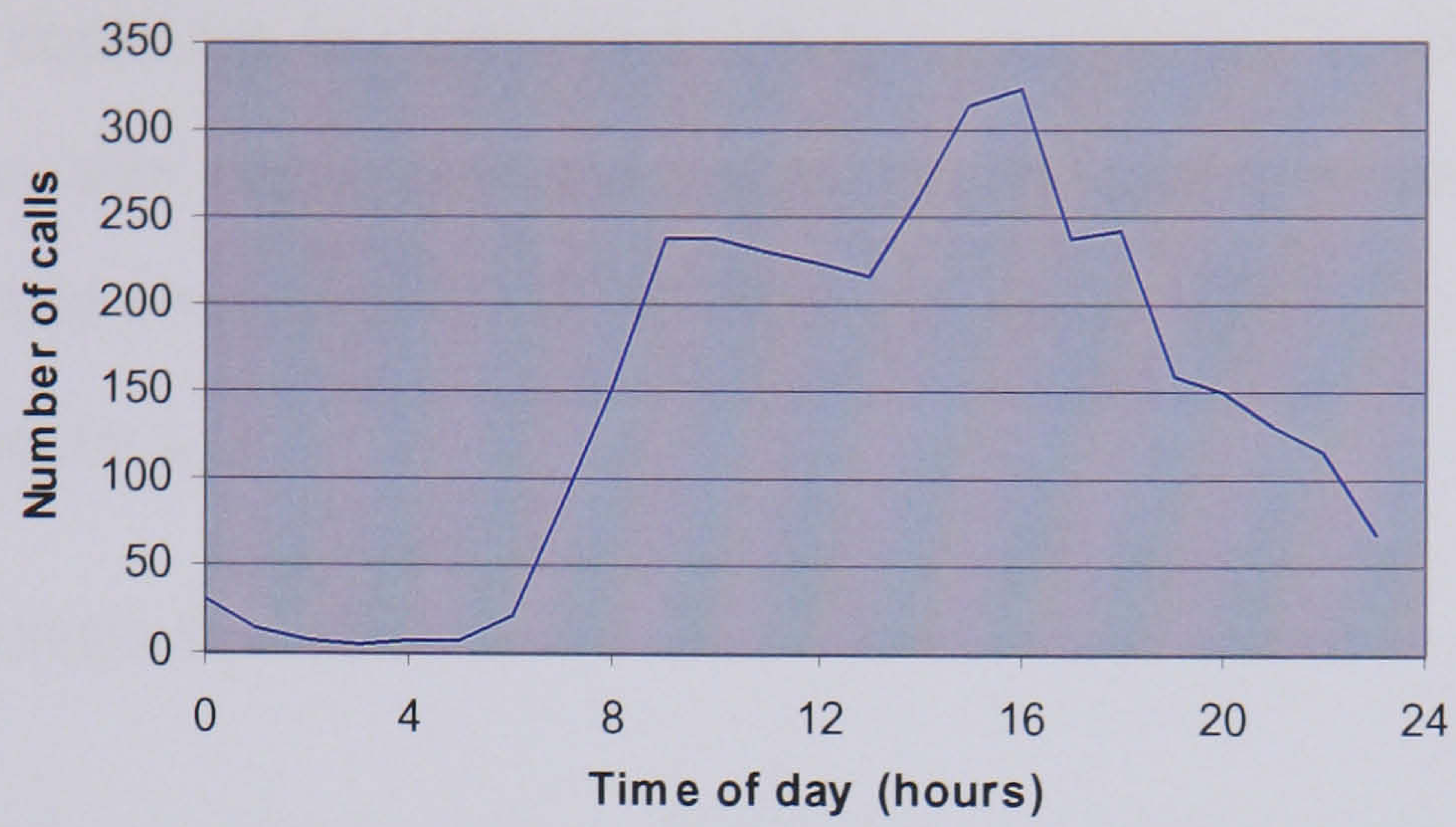

Figure 4-2[28] *Demand as a function of the time of the day*

In the following sections three additional effects will be discussed and the mathematical model modified to give the desired behaviour.

4.1.2.2.1 The time of day effect.

The first effect that we have not taken into account is the time of day effect. In Figure 4-2 it can clearly be seen that although the price per minute is higher during peak hours the demand is also higher. This is due to the fact that more business users use the phone and they have lower elasticity of demand than private users. In addition, people are more mobile and can have less access to fixed line telephones. At the end of the business day the demand drops significantly although the price is also significantly lower, because the proportion of private users (with higher demand elasticity) accessing the network increases.

In the context of this thesis, the time of day effect will be defined as the expected number of users that will attempt to access the network at any given time of the day. It will manifest itself as significantly higher proportion of people attempting to access the network during peak hours (irrespective of the higher

[28] Data courtesy of BT.

price) and could be implemented using a polynomial function, for example. However, in this case, the change in the expected number of calls as a function of the time of day will be taken as equivalent to the number of calls shown in Figure 4-2.

4.1.2.2.2 Substitution effect.

Although the time of day effect can account for the drop in demand during off-peak hours, this contradicts the empirical evidence from UK's T-Mobile, who experienced a significant increase in off-peak number of calls when making their calls free [11]. This suggests that there is another powerful effect that has not been taken into account so far. One possible explanation for this rapid increase in demand can be the fact that most mobile users have fixed line telephones at home that offer cheaper calls. As soon as the price of mobile calls becomes lower than the fixed line telephone, as was the case with free T-Mobile calls, the demand can increase dramatically[29].

The impact of the substitution effect on the demand function can be corrected by the addition of an extra term (4-8) to represent this sharp increase in demand at low prices.

$$Q_{corrected} = \left(Q + EP_Q^{-\beta}\right) \qquad (4\text{-}8)$$

E - correction coefficient;

The substitution effect would be active only when the price in the network falls below a threshold cut-off price P_{sub}. In the case of the cellular networks this would be the price of the price of the fixed line calls. Therefore, the

[29] Fixed lines still carry a significant proportion of call minutes and in 1998 fixed line networks in the UK carried 142,897 million call minutes, compared to 9,542 million carried by cellular networks: a difference of 1,398% [3].

corrected demand function becomes (4-9), plotted in Figure 4-3 for $A = 100$, $E = 10$ and $P \in [0.01, 0.5]$.

$$Q = \begin{cases} A(t)e^{-\beta P_Q} & P_Q \in [P_{sub}, P_{max}] \\ \left(A(t)e^{-\beta P_Q} + E P_Q^{-\beta} \right) & P_Q \in [P_{min}, P_{sub}) \end{cases} \qquad (4\text{-}9)$$

P_{sub} - Substitution effect cut-off price.

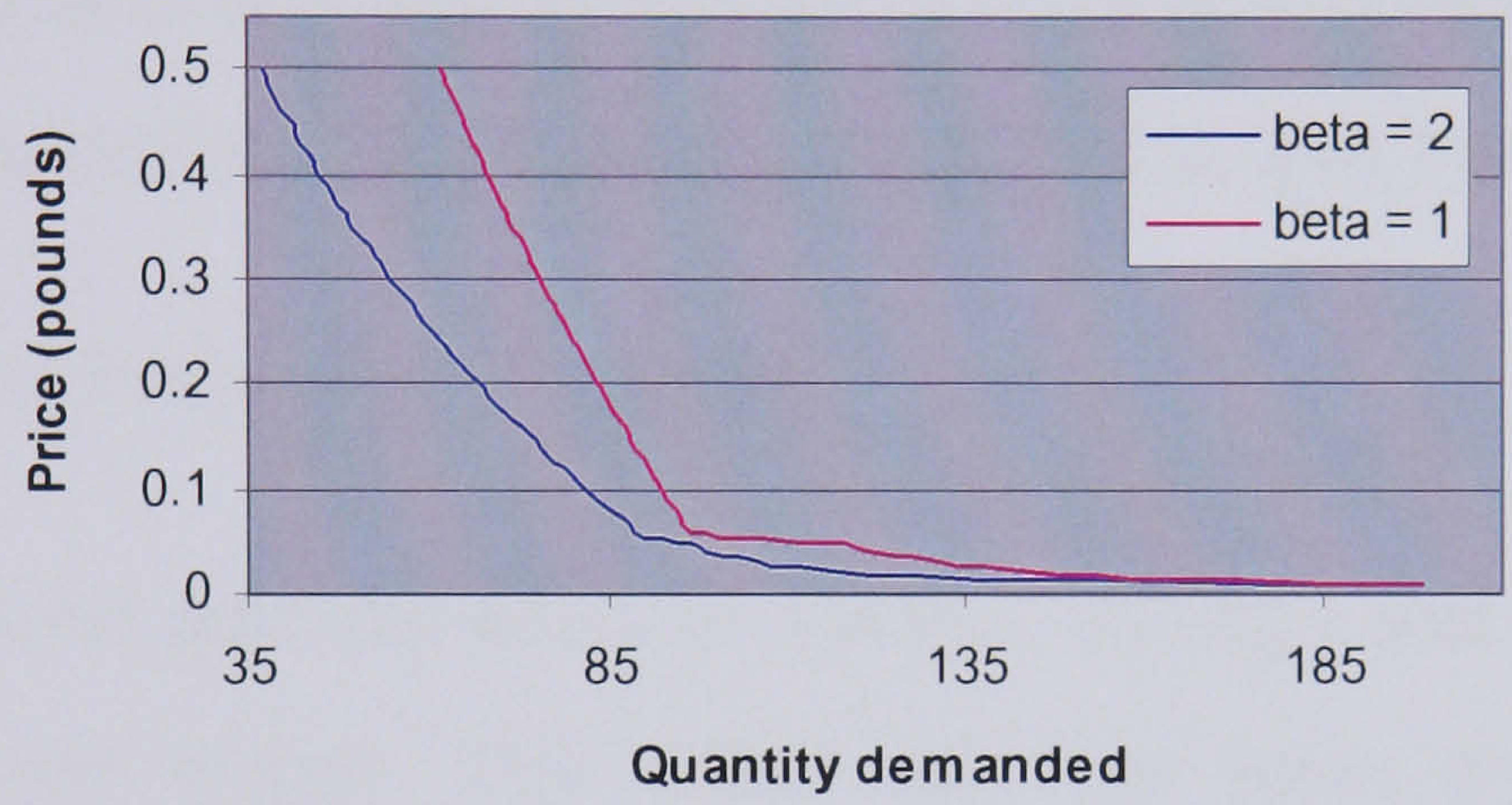

Figure 4-3 *Corrected exponential demand function for various price elasticity* β

4.1.2.2.3 Existing pricing bias.

Finally, the actual data to which we compare the output of the model is "biased" due to pricing information already present in the system. To offset this bias, the difference between the new dynamic price and the "inherent" or existing price will be taken, as a more accurate measure of peoples' willingness to pay. As a result, when the price is equal to either the peak or off-peak price already present in the call data, the expected number of calls from the model will be equal to the number of calls in the raw data. If the dynamic price is lower than the existing price, a higher proportion of people will attempt to access the network and vice-versa. This behaviour is particularly easy to implement with the exponential function of equation (4-6).

$$Q = \begin{cases} A(t)e^{-\beta(P_{dynamic}-P_{bias})} & P \in [P_{sub}, P_{max}] \\ A(t)e^{-\beta(P_{dynamic}-P_{bias})} + (P_{dynamic})^{-\beta} & P \in [P_{min}, P_{sub}) \end{cases} \quad (4\text{-}10)$$

$P_{dynamic}$ - The dynamic price at time t

P_{bias} - Inherent peak or off peak price.

To summarise, a modified exponential demand function will be used to represent the compound demand curve for cellular services. Taking into account the time of day and substitution effects and allowing for a pricing bias in our raw data the simplified demand function changes from (4-6) to (4-10).

4.1.3 Effect of Price on Call Length.

The second potential effect of dynamic pricing could be the reduction or increase in call length. This effect has not been considered by other researchers, but results from the user survey conducted as part of this project indicate that a significant proportion of people consciously reduce the duration of their calls during peak hours when the price is higher. Surprisingly, as noted earlier this is true even for people who do not pay for their calls (see Appendix C). Therefore, we can expect that a customer, confronted by a high price and choosing to proceed with making the call, will reduce the duration of their call in an attempt to reduce its cost. Although this particular aspect of user behaviour can have a very significant effect on the impact of dynamic pricing, its accurate modelling would require access to very sensitive data on current user behaviour. Attempts to access this data proved unsuccessful and, therefore, it was decided that the effect of call length would not be studied further, as the mathematical model would have been purely speculative.

4.1.4 Effect of Price on User Mobility.

The main difference between fixed and wireless networks is that in cellular networks users can, to a certain extent, choose the location in which to initiate a call. Therefore, in order to predict the effect of dynamic pricing accurately this spatial freedom has to be taken into account.

Following assumption (3), only mobility within the system and, in particular, only mobility due to changes in price will be considered. As users act rationally it is reasonable to assume that they will tend to move to areas or cells with lower prices and avoid expensive areas. This is equivalent to a situation in which users look at the price of the calls in their area (cell) and, on finding it too expensive, decide to delay the call until they find a cheaper area or move to an alternative location. Although this scenario does not seem very realistic at this point in time, the potential advances in location based services even in the very near future will enable users to know exactly where they are in the network. This assertion will be discussed further in section 4.1.4.5 below.

Various mathematical models have been suggested in the literature for describing user mobility in a cellular system. The advantages and disadvantages of three such modes, the gravity, the fluid and the Markov will be compared, and the most suitable model developed to represent mobility in a dynamically priced network.

4.1.4.1 Gravity model of trip distribution.

The gravity model developed in transportation studies can be used to predict the most probable distribution of number of trips between two regions depending on the distance between the regions [7], [81]. It is defined as:

$$T_{ij} = A_i B_j O_i D_j f(d_{ij}) \tag{4-11}$$

$$A_i = \left[\sum_j B_j D_j f(d_{ij}) \right]^{-1} \tag{4-12}$$

$$B_j = \left[\sum_i A_i O_i f(d_{ij}) \right]^{-1} \tag{4-13}$$

T_{ij} - Total number of trips to a zone;

d_{ij} - Distance between zones i and j;

O_i - Total number of trips originating in i;

D_j - Total number of trip destinations in j.

The additional constraints A_i and B_j, also referred to as balancing factors, are added to ensure that, if the number of origins or destinations in a particular zone are doubled, the number of trips between these zones will not quadruple.

This model represents aggregate traffic and is particularly accurate for modelling traffic volumes between two towns or cities [7]. An advantage of this model is that it can be used to describe the most probable distribution of trips between two zones as a function some impedance factor, such as the cost of the trips. However, it requires the calculation of many parameters, which makes simultaneous modelling of many locations very difficult.

4.1.4.2 Fluid mobility model.

The fluid models view traffic flow as the flow of fluid, and use the following definition:

$$N = \rho \pi v L \qquad (4\text{-}14)$$

N - Number of site crossings per minute (circular region);

ρ - Population density;

v - Average speed of the nodes;

L - Region diameter.

This model also describes aggregate traffic and is very suitable for describing the traffic volume between two cities or countries, for example. Its limitation stems from the fact that "... since average population density and average velocity are used, this model is more accurate for regions containing a large population." (Lam *et. al.*, 1997: 82) [7].

4.1.4.3 Markov Model.

The Markov (also known as *random walk model*) is the third type of model used to generate patterns for user movements. In contrast to the *gravity* and the *fluid* models, it describes individual movement behaviour. In this model, the subscriber behaviour is described through a matrix or site topology model, which states the probability of the user moving to an adjacent area or cell. This is very similar to having a pedestrian who decides which way to turn (left or right) while walking in a town. Examples of probabilities for a subscriber moving to any of the adjacent cells are shown in Table 4-1.

Future cell / Present cell	Cell 1	Cell 2	Cell 3
Cell 1	0.5	0.1	0.4
Cell 2	0.4	0.2	0.4
Cell 3	0.1	0.3	0.2

Table 4-1 *Probability movement table for a subscriber*

If a user is in cell 1, for example, there is a 50% probability that they will remain in cell 1, a 10% probability of their moving to cell 2 and a 40% probability of their moving to cell 3.

Although this model describes movement of individual users in a location area [7], it does so in a random manner and is not very suitable for describing trips.

From the three models discussed above, it can be seen that the gravity model is the most suitable for the description of user mobility as a function of price. The model will be modified to represent the expected user behaviour and simplified, using assumption (4) on page 99, for easier manipulation.

4.1.4.4 Simulation mobility model.

The suitability of the gravity model to determine the most probable distribution of the number of trips between two regions depending on the attractiveness of the destinations has been discussed by Wilson [81]. The gravity model will be adapted to predict the number of calls generated between two cells[30]. The analogous call gravity model is defined as:

$$\Psi_{ij} = \kappa O_i D_j \left(1 / d_{ij}^2 \right) \qquad (4\text{-}15)$$

κ - A constant;

Ψ_{ij} - The total number of calls between cells i and j;

$\frac{1}{d_{ij}^2}$ - Impedance factor;

O_i - Total number of users in cell i;

[30] The constraint that there will be no mobile to mobile calls will be imposed after the model has been developed.

D_j - Total number of users in cell j.

Similar to the original gravity model, this formula has one major deficiency: If the number of users between originating and destination cells doubles, the number of expected calls between those cells will quadruple. Therefore, to correct this deficiency, additional constraints have to be imposed:

$$\sum_i \Psi_{ij} = D_j \qquad (4\text{-}16)$$

$$\sum_j \Psi_{ij} = O_i \qquad (4\text{-}17)$$

To satisfy these constraints, the corrective constants (balancing factors) A_i and B_j are again introduced.

In this case, the impedance function also has to be changed (see below for explanation of changes) and $1/d_{ij}^2$ becomes:

$$\frac{1}{d_{ij}^2} = e^{-\alpha\left(P_{dynamic} - P_{bias}\right)} \qquad (4\text{-}18)$$

α - Location mobility elasticity of users;

P_{bias} - Inherent peak or off peak price;

$P_{dynamic}$ - Dynamic price in cell i.

The degree of customer reaction to the changes in price in the region will be represented by the parameter α. This parameter represents the location price sensitivity of the users. It can take both negative and positive values and can be used for calibration of the model as suggested by Evans [82].

A correctly calibrated gravity model would give correct predictions for the movement of the mobile users (Bouchard and Pyers [83]). If we define the mean cost of a call in the network as the averaged price all users in the network pay at any given time, the increase in α has the effect of reducing the mean cost of calls. This is equivalent to a larger proportion of users moving to cheaper cells to make calls. Therefore, α can be seen as the users' propensity or willingness to move.

The price factor P_{bias} is introduced, as before, to offset the pricing bias present in the current network. The user mobility model thus becomes:

$$\psi_{ij} = A_i B_j O_i D_j e^{-\alpha(P_{dynamic} - P_{bias})} \qquad (4\text{-}19)$$

$$A_i = \frac{1}{\sum_j B_j D_j e^{-\alpha(P_{dynamic} - P_{bias})}} \qquad (4\text{-}20)$$

$$B_j = \frac{1}{\sum_i A_i O_i e^{-\alpha(P_{dynamic} - P_{bias})}} \qquad (4\text{-}21)$$

Although this model describes fully the most probable distribution of mobile users, depending on the expected number of users in each cell, when the effect of mobility due to price is taken into account [81], a problem with this model is that in order to find A_i and B_j we need to solve equations (4-20) and (4-21). This can only be done by using an iterative minimisation algorithm. One possible implementation of such an algorithm, using Matlab, can be seen in Appendix D.

However, the model can be simplified by using assumption 4 in section 4.1.1, which stated that we shall only consider mobile originating calls to the fixed system and ignore all calls from the fixed network to mobile terminals and mobile-to-mobile calls. Then the fixed network can be represented, as a cell in which the number of expected users is infinitely large. Therefore, the total number of calls between any cell in the cellular network and the mega-cell comprising the fixed network will be limited only by the expected number of users in the cellular network cell. As a result, constraint (4-20) can be dropped without any loss of generality. Thus, by imposing:

$$\forall A_i = 1 \qquad (4\text{-}22)$$

the model becomes:

$$\psi_{i1} = B_1 O_i D_1 e^{-\alpha\left(P_{dynamic} - P_{bias}\right)} \qquad (4\text{-}23)$$

$$B_1 = \frac{1}{\sum_i O_i e^{-\alpha\left(P_{dynamic} - P_{bias}\right)}} \qquad (4\text{-}24)$$

With

$$D_1 = \sum_i O_i \qquad (4\text{-}25)$$

These equations are much simpler to solve, without the need for iterative calculations. Therefore, this model will be used to represent the mobility behaviour of the users and, through the variation of the mobility elasticity α, various scenarios can be tested.

4.1.4.5 Mobility model constraints.

The original gravity model was developed to represent the most likely distribution of trips in an area and was modified to represent the expected number of calls. This was done on the basis that there are certain similarities between calls and journeys and, therefore, the mobility of users will be affected in a similar way.

Telephone calls and train journeys can be considered similar in the following respects:

1. They are both services and, therefore, not tangible;
2. The supply of the services depends on certain equipment being available to the subscriber, *e.g.* a mobile handset, a ticket or a car;
3. Current charges for the services are dependent on both time of day and destination;
4. They are supplied in real time and affected by real time and random events, such as equipment breakdown, for example;
5. Making a journey requires certain preparation and, so does moving around the network to find a cheaper cell.
6. The provision of constant Quality of Service is difficult and relies on anticipation and back-up equipment.

There are, however, a number of differences:

1. A successful call will depend on the caller and the calling party being available at the same time, otherwise the connection may not happen or the call length will be reduced (for *e.g.* leaving a message);
2. The length of a journey is fixed and paid for in advance. Once the customer has selected the destination, it is difficult to change it, while the duration of a call can be varied depending on the price that is charged; a short and very brief call can be sufficient if the line is expensive.

The similarities identified above, especially (5), suggest that users of transportation services and users of cellular networks could behave in a similar manner. The differences do not affect the mobility of users and so are unlikely to affect the appropriateness of the model. However, there is a particular aspect of the model that could affect its suitability. The model assumes that the user "knows" the price of all possible trips before embarking on the selected one. It is difficult to justify the availability of this "perfect information" for the entire cellular network before a user chooses whether to make a call or not. However, with the advance of location based services it will be possible for users to know where they are and to request information about the prices of the cells around them. This trend is indicated by the latest development in WAP phone services, which allow users dynamic access to on-line information on the Internet. For example, a user can specify a location or a facility in a location they are interested in and a map of the area showing the facility can appear on their screen. At the moment users have to explicitly specify the area they want to call. In the future, with the integration of cellular networks and GPS (Global Positioning System), the location of the mobile user will be calculated automatically, so the user will just have to enter the kind of facility they are interested in. By analogy, the user will be able to request information about prices not only in his/her own but also in all neighbouring cells, as well as instructions how to find the location of the desired cheaper cell. Such future enhancement of the WAP based services is very feasible and there exists an incentive for the service provider to provide this type of information, as it will enable mobile users to take better decisions[31].

[31] It is likely that users will also move to cells with different prices unintentionally, in the course of their normal activities. To decrease the discrepancy between the call starting price and the price within the current cell, call charges should be updated periodically, after a pre-set fixed charge period.

4.1.5 Complete User Behaviour Model.

The complete user behaviour model takes into account both the price sensitivity of demand and the price sensitivity of the mobility of users. Therefore, the formula, which will be tested in a simulation for various parameter values will be:

$$Q = \begin{cases} A(t)B_1 O_i D_1 e^{-(\alpha+\beta)(P_{dynamic}-P_{bias})} & P \in [P_{sub}, P_{tr} \\ \left\{ \begin{array}{l} \left(A(t)e^{-\beta(P_{dynamic}-P_{bias})} + E(P_{dynamic})^{-\beta} \right) \times \\ \left(B_1 O_i D_1 e^{-\alpha(P_{dynamic}-P_{bias})} \right) \end{array} \right\} & P \in [P_{\min}, P_s \end{cases} \qquad (4\text{-}26)$$

$$B_1 = \frac{1}{\sum_i O_i e^{-\alpha(P_{dynamic}-P_{bias})}} \qquad (4\text{-}27)$$

4.2 Price Setting Methodologies.

After deciding on the charging units and the type of pricing strategy to be used in a cellular communication network, the main issue facing the network operators is the choice of monetary value per charging unit. According to classic economics theory this choice will depend on both the objective of the network operator and the elasticity of demand.

4.2.1 Classic Optimal Price Setting Methodology.

Economic theory postulates that the optimal pricing methodology that should be used by all firms, independent of the type of market they are in is marginal cost pricing (Varian [79]). Marginal cost prices are calculated by equating the marginal demand for the service to the marginal cost of providing the service. On the one hand, this ensures that users pay the actual cost for providing the service and this will maximise consumer welfare [84], [85]. On the other hand, marginal cost prices ensure that the supplier produces at minimum cost. Marginal cost prices are Pareto efficient *i.e.* they are the optimal prices for both users and suppliers and there are no other prices that will increase the welfare of both users and the suppliers. As pointed out by Anania and Solomon [86] this fact explains the zeal with which industry regulators promote marginal cost pricing.

Marginal cost pricing is attained naturally in a market with perfect competition. By definition this is market in which the amount of goods produced by individual suppliers is negligible, compared to the total amount of goods in the market and, therefore, individual suppliers have no control over the price of the goods. They have to act as price-takers and accept the market price determined by the equilibrium of supply and demand.

However, the telecommunications market is not a market with perfect competition. The proliferation of pricing tariffs employed by network operators for voice and data services (see section 3.1.1.2) suggest that the market is in effect competitive monopolistic or ologopolistic, with individual network operators having some control over the prices they set. Ignoring this condition and setting prices at marginal cost level will put a network operator at a disadvantage. In addition, in telecommunications networks, the marginal costs are zero, except when the network is congested [58]. Therefore, by using marginal cost pricing, the network operator can expect to recover costs only when the network is congested. As a result, a pricing scheme based on pure marginal cost pricing may not be sufficient for recovering the set up cost of communication network operators (Karsten et. *al.* [87], or Wang et. al. [88]).

4.2.2 Optimal Pricing Methodologies for Communication Networks.

The rapid growth in Internet traffic due to its use by private individuals and companies has led to a rapid increase in the demand for network bandwidth. As a result, congestion and delays have become a characteristic feature of the Internet and the users perceive them as negative externalities [79]. Therefore, the control and reduction of congestion has been identified as a paramount issue and is the subject of intense academic research, in particular into the potential use of dynamic pricing for this purpose. In the following sections we will look at suggested dynamic pricing strategies for attaining optimal prices in fixed line communication networks and identify some of their drawbacks.

4.2.2.1 Auction (smart market) pricing.

MacKie-Mason and Varian [89] developed an approach called the "smart market" in which the users bid for the available network bandwidth by stating how much they are prepared to pay for the transmission of their packets. The network collects all the bids in a given time interval and arranges them in descending order by price. The price for the time interval is taken to be the bid for the last packet that will utilise the available capacity. The admitted packets that will utilise the network fully are then delivered, while the leftover packets are either returned to the users or re-routed to a slower network. Therefore, customers do not pay what they bid but the market clearance price, which is always the lowest of the prices of all admitted packets. The final accepted bid price, which is taken as the actual price for the bandwidth, is achieved using the classical demand-supply equilibrium. It represents the marginal willingness of consumers to pay and it is shown by MacKie-Mason and Varian [90] that the congestion revenue equals the optimal investment in capacity expansion. A particular problem with this approach is that the highest bid packets are delivered first and this can lead to a backlog of lower bid packets. In addition, although it can be applied as an effective means of congestion control, its application to cellular telephony is limited. The overhead that will be generated over the radio interface as a result of informing users that their packets were not accepted, for example, will be significant and will reduce the throughput of the network. In addition, there is uncertainty as to whether the packets offered by the bidder will be accepted in the following time interval and the total delay the user will encounter as a result cannot be predicted.

4.2.2.2 Shadow pricing.

This approach is based on retrospective pricing on the basis of the congestion and inconvenience caused by the user to other users in the network and was suggested by Gibbens and Kelly [84]. A "shadow" price is defined, in an optimisation problem, as the marginal improvement on the optimised function when a single additional unit in one of the constraints is added (Rayes and Min [91]). In communication networks, the shadow price of sending a packet is the congestion cost to other packets, *i.e.* the cost to other users whose packets are dropped as a result of the processing of this packet. It is impossible for the network to calculate this cost at the time the packet is being sent, but it can be estimated by marking the packets. One possible marking strategy, for example, is for the routers to mark all packets that arrive after the buffer is full and continue marking until it is empty again. The charge that a user has to pay is proportional to the number of packets that have been marked during transmission. This mechanism can be used to control congestion: when congestion arises, prices go up and, therefore, users are inclined to reduce their sending rate. A drawback of this pricing scheme is that its application relies on the congested router marking all packets that it has forwarded and which have caused congestion. Therefore, it has to identify the start and finish of the flow that has caused congestion. However, the self-similarity[32] of Internet traffic ([92],[93]) makes it very difficult for routers to identify the start and finish of the flow that caused the congestion and this makes marking the correct packets challenging.

4.2.2.3 Dynamic bandwidth allocation.

John and Liam Murphy [85] suggest an alternative pricing method, by linking marginal prices to the available network capacity, for optimising the benefit to users and controlling congestion. In their approach, users are assumed to act in their own best interest and each user is associated with a personal benefit function which shows the value users place on the amount of bandwidth purchased. A user is assumed to place most benefit on the first few units of bandwidth received and the benefit function levels off as more bandwidth is received. In the economics literature, this behaviour is known as diminishing marginal utility. The benefit function is used in conjunction with the price set by the network to determine the user demand for bandwidth at any time interval. The network operator, on the other hand, has to choose the bandwidth allocation to maximise user benefit while at the same time satisfying the capacity constraints of the system, such as buffer space or channel availability, for example. As a result, optimal prices can be derived which ensure that users will self-select their optimal bandwidth demands. An

[32] Self-similarity is defined as the scale invariance property of the traffic *i.e.* the profile of the traffic remains similar independently of the time scale (seconds, minutes or hours, for example) on which it is observed.

alternative approach taking into account both the different bandwidth and QoS user requirements as well as the load in the network at any given time was studied by Gupta *et. al.* [94]. They compared the performance of a fixed network (Internet) with two priority classes under the conditions of free access, flat rate and optimal dynamic pricing strategies. Their results showed that the optimal dynamic pricing strategy performs as well as the free access (no pricing) strategy with small network load. Both strategies offered higher net[33] and customer[34] benefit for users than the flat rate tariff at small network load. As the load in the network increased the optimal dynamic pricing strategy performed significantly better than either the flat rate or the free access strategies. The results held even after allowances were made for imperfect information and deliberate user demand misrepresentation in the network. These results suggest that dynamic pricing can significantly improve the net and customer benefit users derive from a fixed multi-service network.

4.2.2.4 Provider-based pricing approach.

The three pricing approaches outlined above assume that users are aware of the utility brought to them by purchasing a certain amount of bandwidth and will act in their best interest to maximise their benefit. But findings from INDEX (a project studying consumer behaviour), indicate that this assumption is not always true [95]. Taking on the role of an ISP (Internet Service Provider), the researchers tested various flat rate, usage and time based tariffs and observed the usage profile of their customers. They observed that users do not always choose the strategy that will maximise their

---

[33] By definition net benefits = total value – total delay cost.
[34] By definition customer benefit = net benefit – total price.

benefit. Therefore, a pricing strategy attempting to allocate resources in order to maximise the utility of users will not necessarily do so.

Furthermore, the goal of pricing from the provider-oriented point of view is not consumer utility maximisation (see (a), (b), (c) and (d) in section 3.2.2). As outlined in Chapter 3 the pricing objectives of a network operator can be affected by external business considerations and market factors. Therefore, a much more realistic approach to pricing from the service provider's point of view is to consider the consumer's willingness to pay and attempt to maximise the objectives of the network operator. The factors the network operators can attempt to maximise include:

1. Profit – the difference between the revenue and the short and long term costs;
2. Revenue – generated by calls and other services provided by the network;
3. Utilisation of the available network capacity.

These objectives are critical for the successful operation of network operator in the market place and therefore prices satisfying these will be optimal from network operator's point of view.

4.2.2.4.1 Profit maximisation model.

A very real objective for network operators is profit maximisation. Given expected revenue $R(\xi(t))$, operational and amortisation costs $K(t)$, network capacity C and expected aggregate demand $\xi(t)$, a simple constrained profit maximisation problem for a single communication service over a fixed time period $t \in [0.T]$ is defined as (see [96], [97]):

$$\text{Maximise } \int_{o}^{T} e^{-rt}(R(\xi(t)) - K(t))dt \qquad (4\text{-}28)$$

$$\text{subject to } \xi(t) \leq C \qquad (4\text{-}29)$$

r - Compound interest rate;

$R(\xi(t))$ - Total revenue generated in time interval;

$\xi(t)$ - Aggregate demand;

C - Total network capacity;

$K(t)$ - Operational and amortisation costs.

The generated profit is discounted by the compound interest rate to take into account the fact that future profits are worth less then current profit, as current profit can be invested and bring additional revenue.

Solutions to this model will define the optimal prices that the network operator has to charge in order to maximise their profit. However, this particular model is most suitable for modelling of profit maximisation in the long term, say months or years and not suitable for the telecommunication industry. This is a very specific problem to the telecommunication industry because in the medium and the short term the operational and amortisation costs for network operators are constant, as the main investment of the network operator is in setting up the network infrastructure. Therefore, in the short term, the profit maximisation problem simplifies to a revenue maximisation problem for the network providers. Equation (4-28), on the other hand, can be used by network operators to determine the minimum amount of revenue they have to generate per year, in order to recuperate their investment costs, for example.

4.2.2.4.2 Revenue maximisation model.

The revenue maximisation approach is very similar to the yield management pricing approach adopted by airlines as discussed in section

3.2.6.1.2 above, but can be done on a shorter time scale. The constrained revenue maximisation problem, for a single connection oriented telecommunication service is:

$$\text{Maximise } \int_{o}^{T} e^{-rt} R(\xi(t))dt = \int_{o}^{T} e^{-rt} \xi(t) P(t) \tau dt \qquad (4\text{-}30)$$

$$\text{subject to } \xi(t) \leq C \qquad (4\text{-}29)$$

r - Compound interest rate;

$R(\xi(t))$ - Total revenue generated in time interval;

$\xi(t)$ - Aggregate demand;

$P(t)$- Price at time t;

C - Total capacity of the network;

τ - Line/call holding time.

Karsten *et.al.* [87] use a variation of this model to find the revenue maximising prices for an integrated multi-service digital fixed network and use it to derive a very comprehensive pricing model. However, the pricing function they suggest is extremely complex and difficult to derive and its practical application in networks is, therefore, limited.

In addition, the revenue maximisation problem as defined in (4-30) suffers from one major deficiency, namely that, it assumes that network operators can set the prices they charge freely and without any interference. This is only the situation in unregulated monopolistic markets, where the monopolist is the price setter. In the situation of monopolistic demand equation (4-30) can be solved fairly easily for many types of demand and a result from classic economics theory shows that the revenue maximisation price is the one, which makes the elasticity of demand unity [77]. For the exponential demand function (4-10), the price maximising the revenue for the network operator would be[35]:

$$P(t) = \frac{1}{\beta} \tag{4-31}$$

The revenues for expected number of users $A = 78$ and different mobility elasticity β are plotted in Figure 4-4 and it can be seen that the revenue maximising prices are as suggested by the model. The optimal prices for $\beta \in [0.5, 4.0)$, for example, will be $P(t) \in [2.0, 0.25)$ units, for example, and these prices will also satisfy the constraint (4-29) and thus, according to (4-30) , in theory, be acceptable prices for the network operator. However, these prices will not be acceptable in practice as the cellular market is very competitive and the network operators cannot set prices independently. In fact, as pointed out in section 1.1, Figure 1-4, the trend is for reduction in prices, and shows that the cellular network operators have to act to a degree as price takers, in order to be successful. Therefore, the revenue maximisation model suggested by equation (4-30) cannot be used readily for derivation of optimal dynamic prices for network operators.

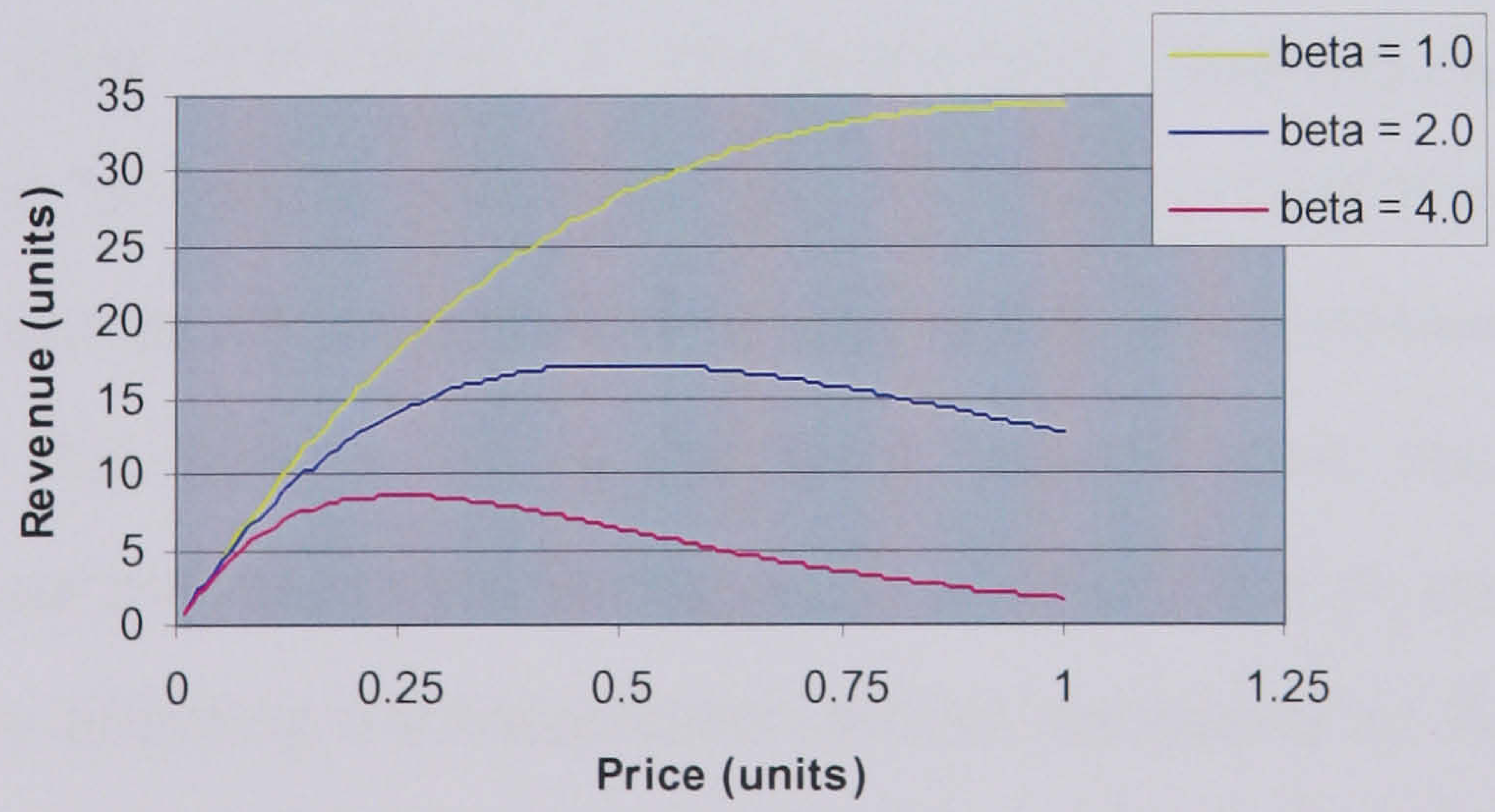

Figure 4-4 *Revenue maximisation pricing*

[35] Ignoring user mobility and substitution effects for simplicity.

This thesis will suggest and evaluate the efficiency of three alternative methodologies for derivation of optimal values for dynamically priced tariffs in a competitive market environment. This will be done in the framework of a scenario in which a hypothetical operator has an existing network, which is up and running. The existing pricing in the network will be the traditional two tariff pricing strategy with peak and off-peak prices. New dynamic pricing tariffs will be introduced and their effect will be evaluated against the performance of the "existing" pricing strategy. In calculating the values of the dynamically priced tariff the network operator will have available at its disposal information regarding the current performance of the network as well as "educated" guesses regarding the behaviour of users and competitors. The tariffs' optimality will be considered from both network providers and users point of view. This is a complex problem that cannot be solved analytically; therefore, a simulation will be used to evaluate the effect of the different tariffs.

4.3 Chapter Summary

This chapter began with a discussion of the expected effect of dynamic pricing on user behaviour. A comprehensive mathematical model for evaluating this effect was introduced which took account of the price elasticity of user demand as well as the existing pricing bias in the network. Substitution and time of day effects due to the fixed network were also modelled. In addition, a user mobility model showing the effect of price on user mobility was developed by adapting a transportation model discussed by Wilson [81]. The final stage in any pricing policy: the setting of monetary units for charging network resource usage by customers was discussed. A literature overview of suggested optimal price setting strategies from an economic and consumer welfare point of view was given and the drawbacks of these approaches highlighted.

Chapter 5

This chapter will present and evaluate the effectiveness of a provider-oriented market driven *ad hoc* dynamic price setting policy, taking into account the minimum and maximum price the network providers want to charge and the average price their competitors in the cellular market charge for calls. A seven-cell OPNET™ simulation model, developed to test the effect of dynamic pricing on cellular networks are presented. Results from the OPNET™ simulation using *ad hoc* competition driven dynamic pricing are shown and the resulting changes in network behaviour analysed. These indicate that the behaviour of the network is very sensitive to the shape of the proposed pricing function.

5.1 Competition Driven Ad Hoc Pricing Model.

Implementation of the dynamic pricing algorithm outlined in section 3.3.1.1 requires the derivation of a dynamic pricing function for mapping the prices to the load in the network. This requires the network operators to make decisions regarding the maximum and minimum price they will charge and the granularity of the intermediate prices. In a competition driven market the network operator has to set these prices bearing in mind the pricing of other competitors. The accurate determination of the minimum price is critical for avoidance of overload in the network, and its value can be determined on the basis of historic information about the network capacity and user demand elasticity. The maximum price also has to be considered carefully, for example, taking into account any limit set or recommended by the industry regulator.

Finally, the network operator has to set the granularity of the intermediate dynamic prices. It is in network provider's interest to determine these intermediate prices bearing in mind the prices set by other operators, to avoid consumer churn. For example, the network operator may want to ensure that the average price in the network is the same, or even below, the fixed price of its competitors. This can be seen as a measure for securing user interest and the network operator's market share. For example, if 70% of prices are below the average market price, the well being of an average user will be higher than if just 30% of the prices are below the average market price.

If we call the minimum and the maximum prices a network operator chooses $P_{\min}$ and $P_{\max}$, the objective is to determine the prices to be charged as the load in the network changes from 0 to $Q_{\max}$ percent. In order to safeguard user interest and to protect themselves from the industry watchdog, a provider may wish to impose an additional constraint to ensure that the intermediate prices they charge are the minimum possible[36]. In mathematical terms this is equivalent to finding the curve $P(q)$, which minimises the area bounded by itself, given a particular divergence of the price curve from the straight line connecting the maximum and minimum price M. The divergence M will depend on the average market price and the positioning of the network operator with respect to it. It can also be defined as a consumer utility or

[36] This assumption is not necessarily in network provider's best interest. The acceptable prices for the industry regulator could be higher.

welfare index because increasing M will lead to users seeing lower intermediate dynamic prices, thus potentially increasing their welfare.

The optimisation problem faced by the network operator is:

$$Minimise \int_{Q_{\min}}^{Q_{\max}} (P_{\min} + P(q))dq, \tag{5-1}$$

subject to

$$\sqrt{1+P(q)'^2}\, dq = M \tag{5-2}$$

and the boundary conditions:

$$P(Q_{\min}) = P_{\min} \tag{5-3}$$

$$P(Q_{\max}) = P_{\max} \tag{5-4}$$

$P_{\min}$ - Minimum price;

$P_{\max}$ - Maximum price;

q - Load in the network;

$P(q)$ - Intermediate prices in the network;

M - Consumer welfare index;

$Q_{\min}$ - Minimum load to start dynamic pricing;

$Q_{\max}$ - Maximum load in the network.

Taking $Q_{\min} = 0$ and using calculus of variations, the curve minimising the area bound by itself can be found using Lagrange's multiplier (Butkov [98]). Thus we have to solve:

$$H(P, P', q) = \int_0^{Q\max} \left(P_{\min} + P(q) + \lambda\sqrt{1+P(q)'^2} \right) dq \tag{5-5}$$

As (5-5) does not explicitly depend on q, a special formula for finding the functional can be used (Irving and Mullineux [99]) and the Euler equation which will be solved becomes:

$$P_{\min} + P(q) + \lambda\sqrt{1+P(q)'^2} - \lambda\frac{P(q)'^2}{\sqrt{1+P(q)'^2}} = const = K \qquad (5\text{-}6)$$

The general solution to this optimisation problem is (see Appendix E for detailed solution):

$$P(q) = K - P_{\min} - \sqrt{\lambda^2 - (h-q)^2} \qquad (5\text{-}7)$$

To solve the problem for boundary conditions (5-4) and (5-5) and determine an exact solution the following system of equations has to be solved:

$$\begin{gathered} P(0) = P_{\min} = K - P_{\min} - \sqrt{\lambda^2 - h^2} \\ P(Q_{\max}) = P_{\max} = K - P_{\min} - \sqrt{\lambda^2 - (h-Q_{\max})^2} \\ \int_0^{Q_{\max}} \sqrt{1+P(q)'^2}\, dq = M \end{gathered} \qquad (5\text{-}8)$$

Simplifying and rearranging this (see Appendix E) leads to a system of two non-linear equations. These can be solved numerically (see Appendix D for a suggested Matlab program).

$$\begin{gathered} P_{\max} - P_{\min} = \sqrt{\lambda^2 - h^2} - \sqrt{\lambda^2 - (h-Q_{\max})^2} \\ M = \cos^{-1}\frac{(h-Q_{\max})}{\lambda} - \cos^{-1}\left(\frac{h}{\lambda}\right) \end{gathered} \qquad (5\text{-}9)$$

Then K can be found and the exact pricing function meeting the initial conditions determined.

The calculated price functions derived from equation (5-9) with $P_{\min} = 0.01$ units, $P_{\max} = 0.4$ units, $M = 0.01$ and $M = 1.0$, corresponding to market average prices of 0.22 and 0.13 units respectively is shown in Figure 5-1. Using information on current market prices (see Appendix B) the weighted average price of the market is 0.13 units. Therefore, these pricing functions will offer one average dynamic price, which is similar to the existing price and one, which is higher. As $M \to 0$ the pricing function becomes more linear, and as $M \to Q_{\max} + P_{\max} - P_{\min}$, its upper limit, the function becomes less linear.

This pricing scheme does not take into account the possibility of call blocking *i.e.* demand exceeding the available capacity. In this case an exceptional price of 0.5 units will be charged. In general some degree of call blocking is tolerated, and in this case the exceptional price will be charged when the blocking in the system is above 2%.

The effectiveness of this pricing approach will be tested by simulation for the two values of M used to generate Figure 5-1.

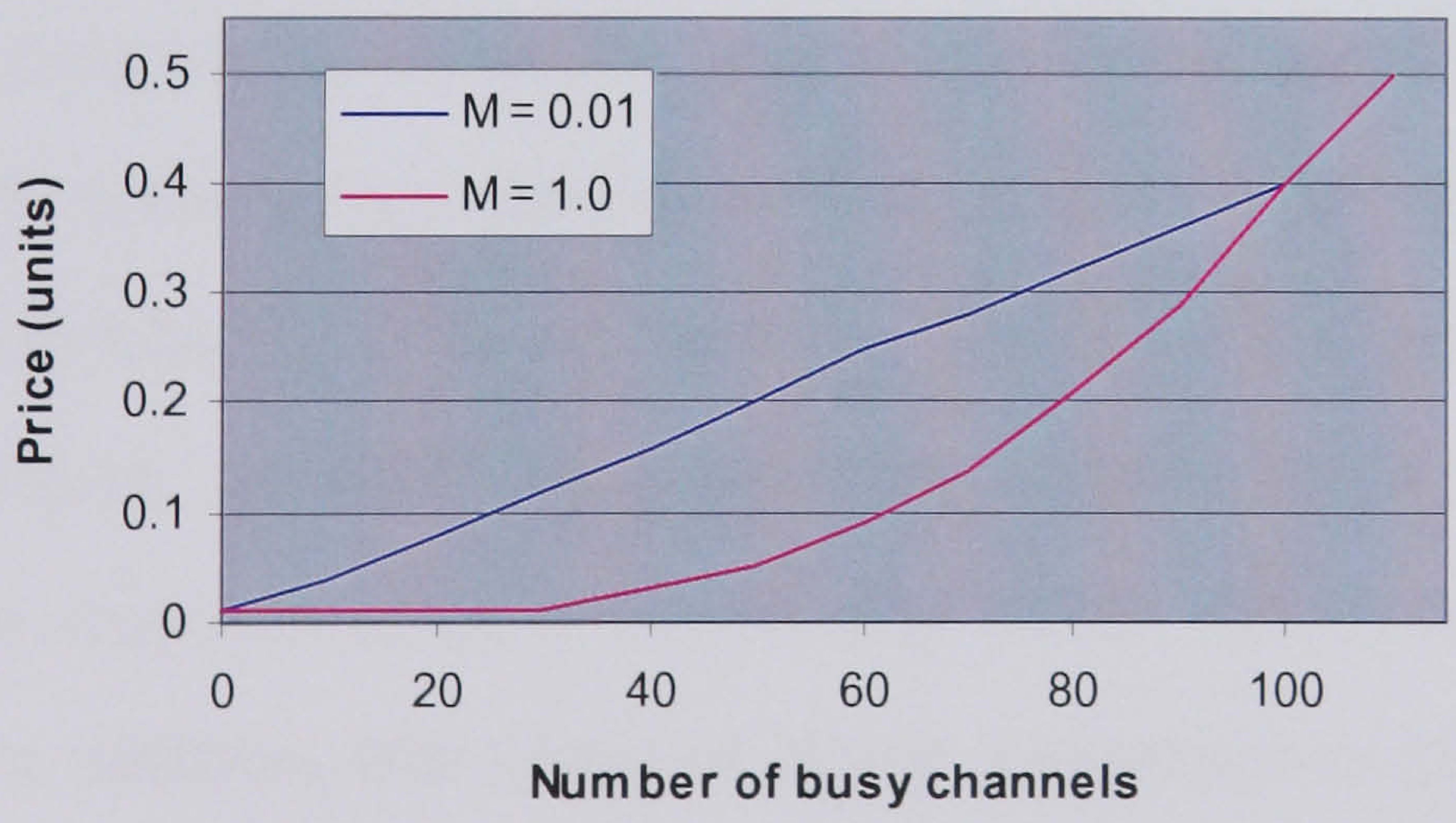

Figure 5-1 *Competition driven ad-hoc pricing function.*

To prevent user confusion by too many prices, the number of possible prices will be decreased by changing the prices only when there is a 10%

change in the available system capacity. So, for example, if the load in the system is between 10% and 20% the price will not change, however, as soon as the load in the network increases to 21%, the price will change. The simulation model developed to test the different pricing strategies is discussed in the following section.

5.2 Simulation Development.

In order to study the change in cellular network behaviour under different dynamic pricing strategies, a simulation model of a seven-cell cluster was developed using the OPNET™ simulation tool [100]. OPNET™ was chosen because of its event driven nature and a comprehensive library of ready-made network and protocol models. Although the ready-made cellular network model implements the functionality of analogue mobile cellular systems as described in the EIA/TIA-533 specification standard, also known as the Advanced Mobile Phone Service (AMPS) [101], rather than a digital one, the template model was a good starting point for the addition of dynamic pricing. However, in order to assess properly its suitability as a model to represent a digital network it is important to answer two questions. First, is the FDMA modulation used in the network significantly different from the TDMA modulation in digital networks [12]? However, as can be seen from section 2.1.3.1, the capacity and operational characteristics of FDMA and TDMA networks are not significantly different. In addition, the mode of signal transmission (physical layer)[37] does not have a direct impact on the implementation of dynamic pricing which affects the system layer[38] of the system, and therefore, for the purposes of this

[37] Layer 1 in the seven-layer OSI model.
[38] Layer 3 in the seven-layer OSI model.

particular research it was not deemed necessary to change the basic framework of the model [102].

The second significant difference between AMPS and GSM is in the structure of the fixed support networks, as the AMPS networks do not have dedicated BSCs in which to implement the dynamic pricing algorithm [26]. However, since the simulated network consists of only seven cells and one cluster, the functionality of the MTSO in AMPS will be equivalent to the functionality of the BSC in a GSM network. Thus, the structure of the original model could be used without any changes.

Extensive changes, however, were made to allow the introduction of pricing information into the model. The generation rate of calls had to be modulated to model the fluctuations in number of calls generated depending on the time of day and the price of the calls in the individual cells. The algorithm for updates in price and generation rate implemented in the simulation is identical to the algorithms described in section 3.3.1.1.

5.2.1 Simulation Model Features.

The main features of the model are listed below:

1. Arbitrary number of mobile units per cell. The total number of mobiles in the network is 80;
2. Arbitrary trajectories for each mobile station. Although available, this particular feature was not used in the simulation, as it was unavailable due to unresolved problems in OPNET™ [100];
3. Voice call set up and call termination modelled explicitly;
4. Call generation distribution for voice and real time calls is Poisson-modulated (see 5.2.2 below for further discussion);

5. Call length distribution is exponential;
6. Price updates occur at user defined intervals. The price update interval used in the simulations is 300 seconds.

The default values of other simulation parameters defined in the user behaviour model equations (4-26) and (4-27), have been summarised in Table 5-1 below.

Parameter	Default Value
E	1/10
P_{bias}	0.05 units – off-peak hours 0.25 units – peak hours
$A(t)$	Historic data from Figure 2-11

Table 5-1 *Other simulation parameters*

In addition to those, at the beginning of a simulation, the user supplies parameters such as the initial call price, total network capacity, and parameters against which the performance of the network is measured, such as, for example, optimal blocking probability.

5.2.2 Mathematical Models for Traffic Generation.

One of the most important aspects of network simulation is the appropriate choice of traffic generation model. Depending on the type of events arriving in the system two types of traffic have been identified: simple and compound.

Simple traffic consists of the single occurrences of discrete events such as, for example, the arrival of telephone calls or ATM cells. It can be described as a point process, consisting of a series of arrival instances $T_i, T_2 ... T_n$ measured from time t_o (Frost and Melamed [103]). So a simple traffic pattern can be defined by two parameters, the number of arrivals in the system N and the inter-arrival times $A_n = T_{n-1} - T_n$, *i.e.*, the time interval separating the last arrival from the previous one. The arrival of voice calls in a GSM cellular network can be modelled by a simple traffic model.

The second type of traffic, on the other hand, consists of batch arrivals and is termed compound traffic. Since at any time unit T_n, there could be more than one unit arriving, in order to describe the traffic fully, one needs to specify a non-zero sequence, $\{\mathbf{B}_n\}_{n=1}^{\infty}$, where B_n is the (random) number of units in a batch. This type of traffic is generated in the network when users browse the Internet, for example, and can be used to model packet based traffic.

5.2.2.1 Modelling of voice and data traffic.

With either type of traffic, the most important aspect of modelling is the determination of the call inter-arrival times. These can be modelled using a number of distribution models depending on the nature of the traffic and an extensive overview is given by Frost and Melamed in [103]. From the traffic models discussed, the most suitable model for representing the arrival of voice calls in the network is the Markov-modulated traffic model. This type of traffic model is relatively simple mathematically and this explains its popularity.

A Markov-modulated traffic model consists of a simple Markov process that, in addition to evolving in time following a Poisson process, is modulated by an additional process. In the simple Markov traffic model, the inter-arrival times A_n are independent and identically distributed. The rationale behind the Markov-modulated type of models is the introduction of an explicit notion of different states in the description of the traffic stream. The additional Markov process is evolving in time and its current state controls (modulates) the probability law of the traffic mechanism.

The most commonly used Markov-modulated process is the Markov-modulated Poisson process, in which the modulation mechanism stipulates that in state k (out of total M) the arrivals occur as a Poisson process with rate λ_k. The generation rate changes with the state of the system. For example, an on-off source can be described by Markov-modulated process: when the system is in state “on” calls are generated at rate λ_{ON} and when the system is in the “off” state no calls are generated.

This type of model will be used to represent call generation in the OPNET™ simulation of the GSM system. Voice call arrival times will be calculated using a Poisson process with rate λ_t, which will depend on the time of day and will evolve as simulation time increases. Call inter-arrival times are presumed to be independent due to lack of information on the correlation factor. The probability of K arrivals in an interval of length t is calculated using:

$$P_K(\lambda_t) = \frac{(\lambda_t)^K}{K!} e^{-\lambda_t} \qquad (5\text{-}10)$$

λ_t - Voice calls arrival rate at time t;

K - Expected number of arrivals.

5.2.3 Simulation Network Architecture.

The network model, as mentioned above, constitutes a seven-cell cluster, which is a common combination in real cellular networks (see Figure 5-2). Within each cell, there are a number of mobile stations and a Base Station (BS) in AMPS or Base Transceiver Station (BTS) in GSM terminology. The seven base stations are linked to a Mobile Telephone Switching Office

(MTSO) in AMPS or the equivalent Base Station Controller (BSC) in GSM. The functions of each of the individual elements are outlined below.

5.2.3.1 Mobile telephone switching office (MTSO)/Base station controller (BSC).

In the model, the MTSO/BSC is responsible for call set up and handoff channel allocation as well as for monitoring lost or dropped calls. This module is also responsible for implementation of the dynamic pricing algorithm. It calculates the price for each indivdual cell based on the information it has on the number of available free channels in the cell and the call blocking probability in the cell.

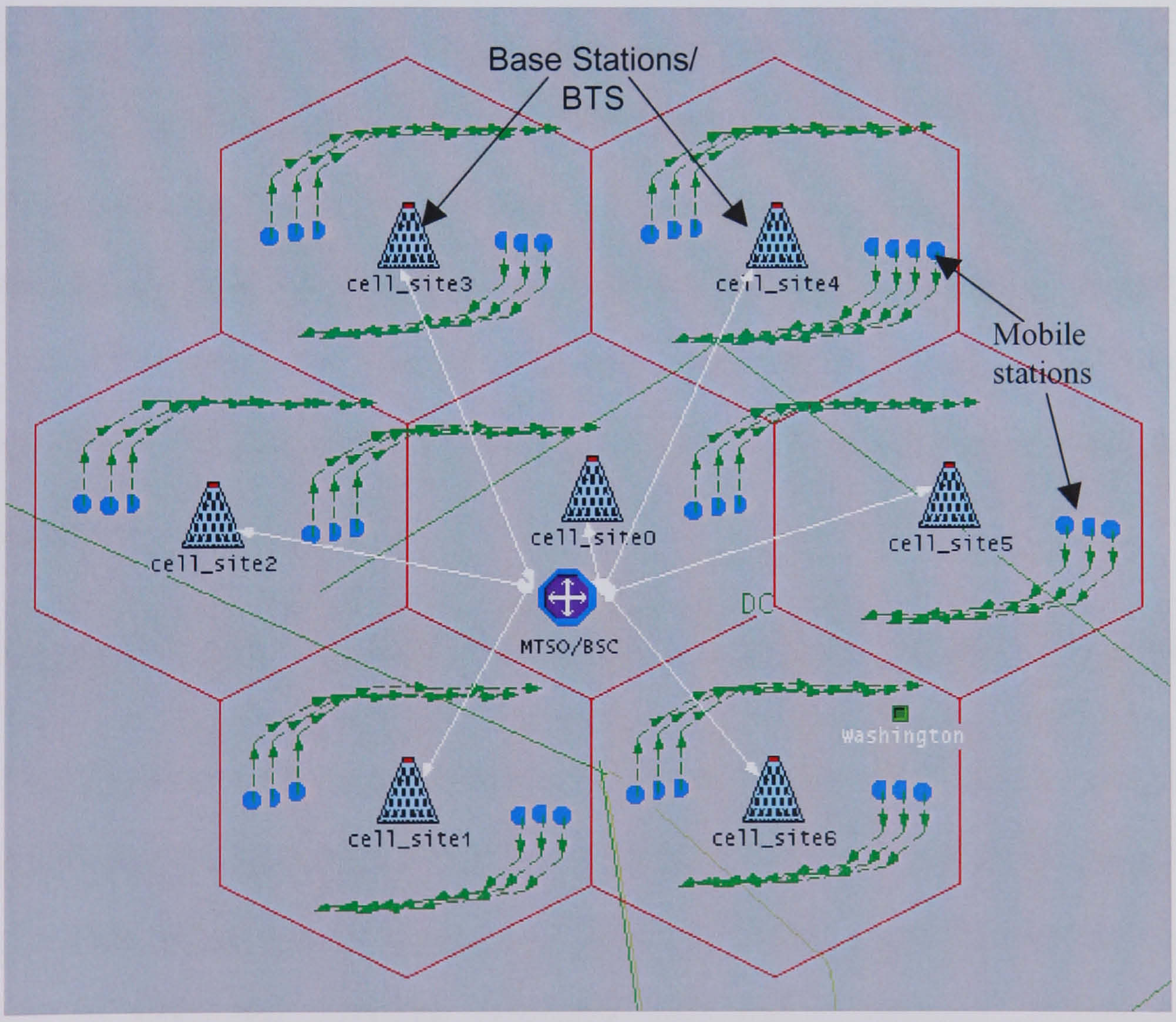

Figure 5-2 *Seven cell network model*

5.2.3.2 Base stations/Base transceiver stations (BTS).

These are responsible for broadcasting the updated prices to the mobile stations.

5.2.3.3 Mobile stations.

In the model, the mobile stations represent the behaviour of the users. Each can be described by its call generation rate, which determines the frequency of calls a user makes at any given time. The value of the generation rate for each mobile station is defined dynamically by the time of day and the price of the calls in the cell, using the demand function described by equations (4-26) and (4-27) above. The generation rate is updated every time the mobile receives a new price from the MTSO/BSC.

The performance of the simulation model was validated by using a fixed flat rate pricing strategy and varying the user inclination to move, denoted by α in the mathematical model, and the quasi elasticity of user demand, denoted by β in the user behaviour model. The results are presented in Appendix F.

5.3 Simulation Results with Competition Driven Ad Hoc Pricing.

The objective of the simulation is to evaluate the effect of the competition market driven pricing function defined in section 5.1 on the performance of the network. This effect will be evaluated in terms of the expected impact on the interests of both the network operator and the customers. Although the network price is allowed to vary, it is bounded within the interval $P \in [0.01, 0.5]$. In order to study the behaviour of the network in the regions of both inelastic and elastic demand, the price elasticity parameter[39] β will be set at $\beta \in [0,3]$. As the elasticity of demand E_d for the exponential demand function is (see section 4.1.2.1)

$$|E_d| = \beta P \tag{5-11}$$

P - Price of the service/product;

this choice of β ensured that $E_d \in [0,1.5]$ and, therefore, inelastic, unit elastic and elastic demand behaviour were tested.

The impact on the interests of the network operator was measured in terms of generated revenue and the efficient utilisation of the available network capacity. In addition, network operators have to ensure that call blocking in the system is kept under control; therefore this statistic was also recorded. The welfare of users, on the other hand, is affected by the total number of successful calls made in the network and the price of the calls. Therefore the effect of dynamic pricing on the number of successful calls and the weighted average price in the network was also monitored for each pricing scenario.

All simulations were run over a 24-hour period and other variables in the system will be fixed as stated in Table 5-1.

[39] Defined in the mathematical model developed in chapter 4 as the sensitivity of users to price changes and termed quasi-elasticity of demand.

5.3.1 Simulation Results with Competition Driven Pricing without User Mobility.

5.3.1.1 Network operator's perspective.

The effect of different dynamic pricing functions on the interests of the network operator was monitored via the revenue generated by the network, the percentage of blocked calls and the distribution of load in the network as a function of time.

Shown in Figure 5-3 is the effect of the linear and non-linear dynamic pricing functions on the total revenue generated in the network. There is a significant difference in the effect of the two pricing functions for all values of price elasticity β. There is a 20% increase in the revenue generated for the network operator with linear dynamic pricing (demand elasticity $\beta > 0.0$), compared to the revenue generated without dynamic pricing ($\beta = 0.0$). The increase remains almost constant for all values of β for which the user demand is inelastic ($\beta < 2.0$).

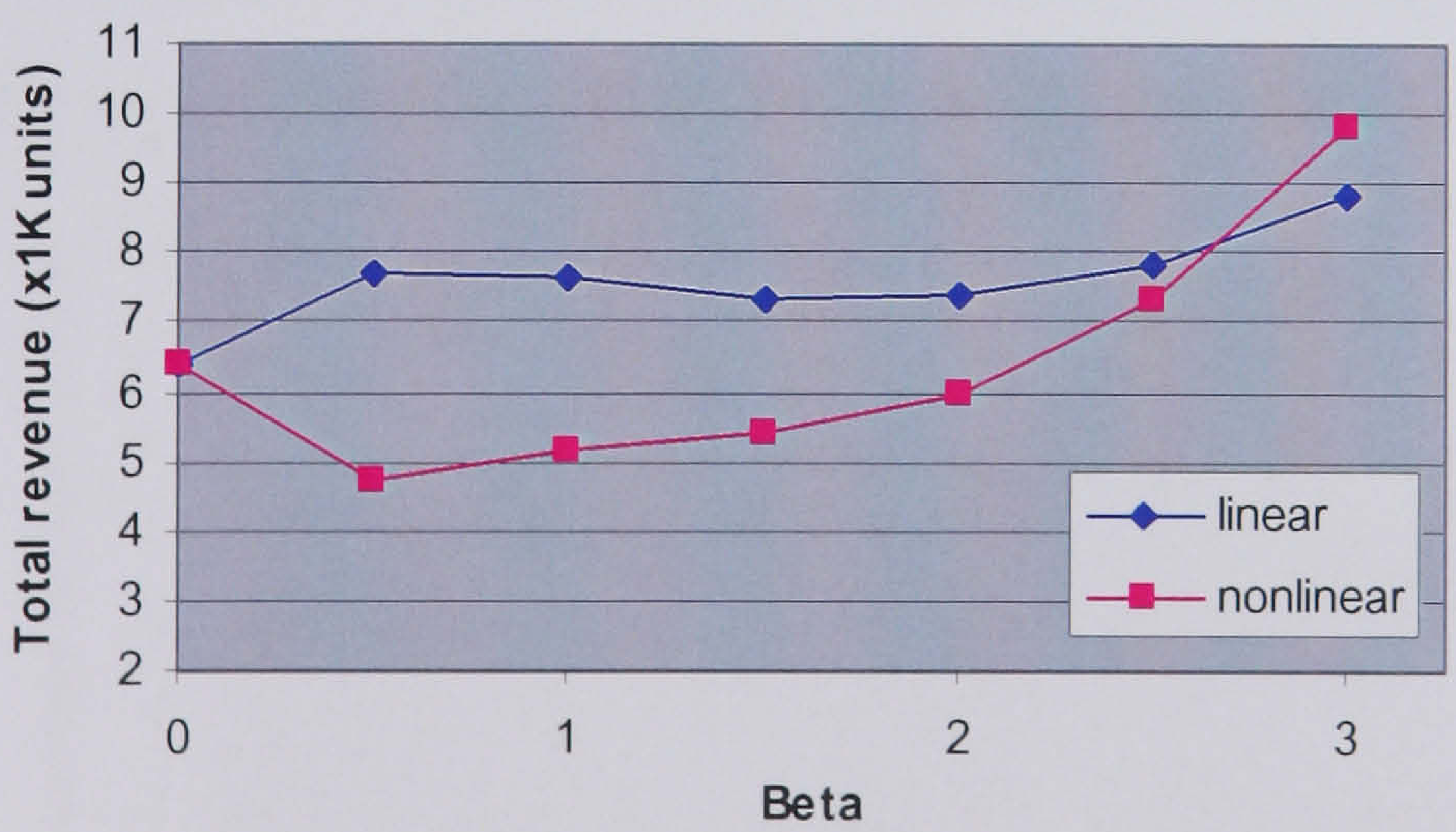

Figure 5-3 *Total revenue generated in the network as a function of* β.

The total revenue generated with the non-linear dynamic pricing function and inelastic demand ($0.0 < \beta < 2.0$) is, on the other hand, less than the revenue generated without dynamic pricing ($\beta = 0.0$). The difference is most significant for $\beta = 0.5$ with a 25% reduction, which gradually reduces as β increases, to a 6% reduction for $\beta = 2.0$.

However, as the demand for services becomes elastic ($\beta > 2.0$) the revenue generated with both pricing functions is higher than the revenue generated without dynamic pricing. In fact, the non-linear pricing function performs better than the linear pricing function for $\beta > 2.5$, generating 13% more revenue. This increase is due to the overall increase in the number of calls during off-peak hours, which will increase the average price in the network and therefore, lead to the generation of additional revenue.

This result highlights the issues that network operators face when attempting to set dynamic prices. Apparently small differences in the pricing function can lead to significantly different results from a network operator's point of view.

This is confirmed by the results for the percentage of blocked calls to total number of successful calls in the network in Figure 5-4.

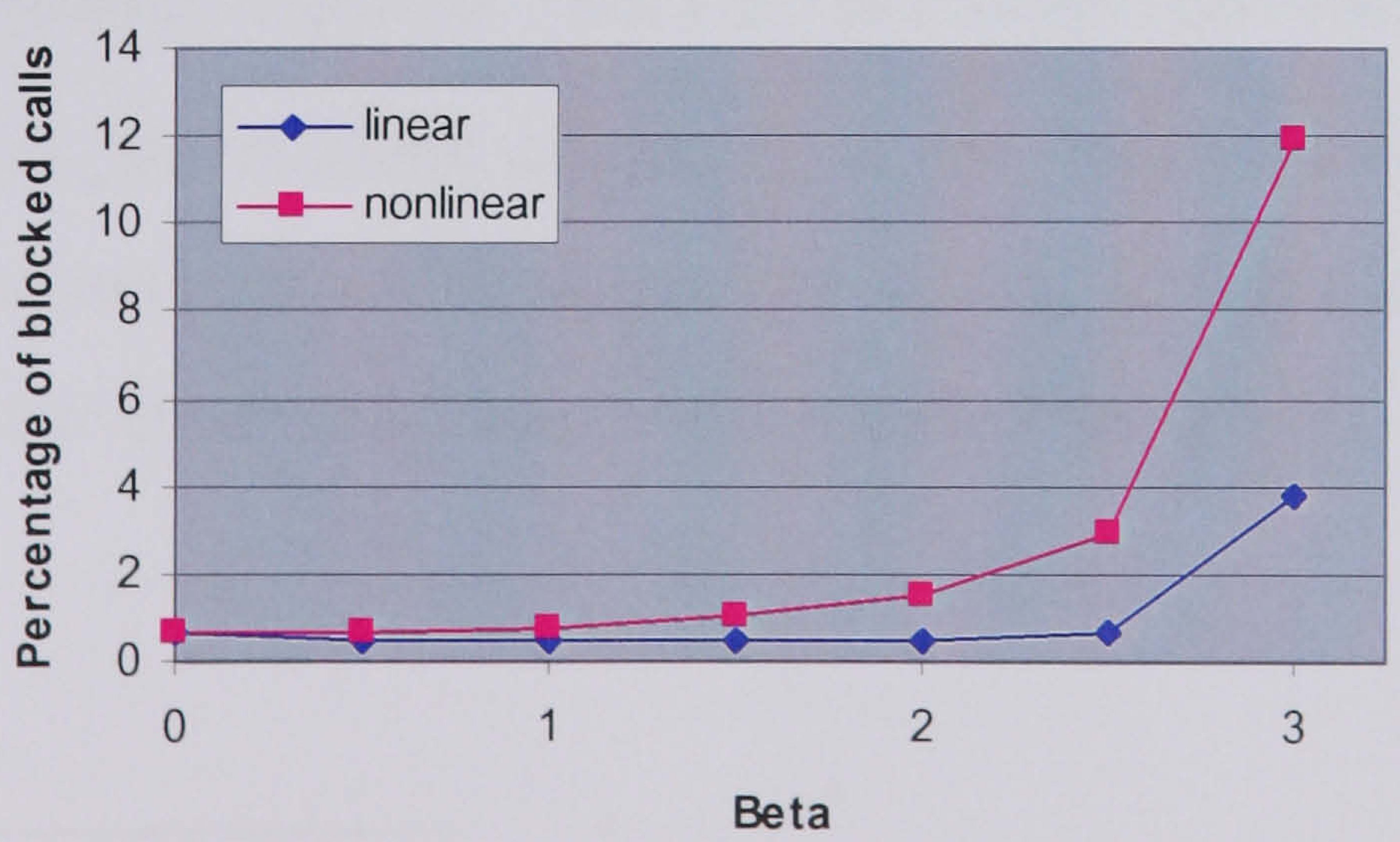

Figure 5-4 *Percentage of blocked calls in the network*

It shows that the linear pricing function offers a significant (up to 25% for $\beta = 2.0$) reduction in the percentage of blocked calls for inelastic demand *i.e.* $\beta \leq 2.0$. However, as demand becomes elastic ($\beta > 2.0$) call blocking abruptly increases (by 9%, compared to blocking without dynamic pricing).

The non-linear pricing function does not offer any improvement in call blocking as the percentage of blocked calls in the network increases as soon as dynamic pricing is introduced ($\forall \beta > 0.0$). As with the linear pricing function the largest increase in percentage of blocked calls occurs when demand becomes elastic ($\beta > 2.0$). This is due to the increased preference of users for off-peak prices, which results in shift of peak hour demand to off-peak hours. Combined with the substitution effect with the fixed network these lead to a significant increase in off-peak demand (see Figure 5-7 below) and a dramatic increase in the percentage of blocked calls (the distribution of the percentage of blocked calls on hourly basis is shown in Figure F-4, Appendix F).

The final parameter of importance to the network operator is the profile of the load in the network over time. This load is defined as traffic intensity in equation (2-10). Traffic intensity also changes as a function of both the price elasticity of demand β and the type of pricing function[40].

For inelastic demand ($\beta \leq 1.0$) the effect on the traffic is not very significant *i.e.* the temporal distribution of the network load remains similar to the network load profile without dynamic pricing. This is true for both the linear and non-linear dynamic pricing functions (see Figure 5-5).

[40] The data presented is averaged over hourly intervals to remove random fluctuations. A sample of the profile of traffic load without this smoothing can be seen in Appendix F (Figure F-10).

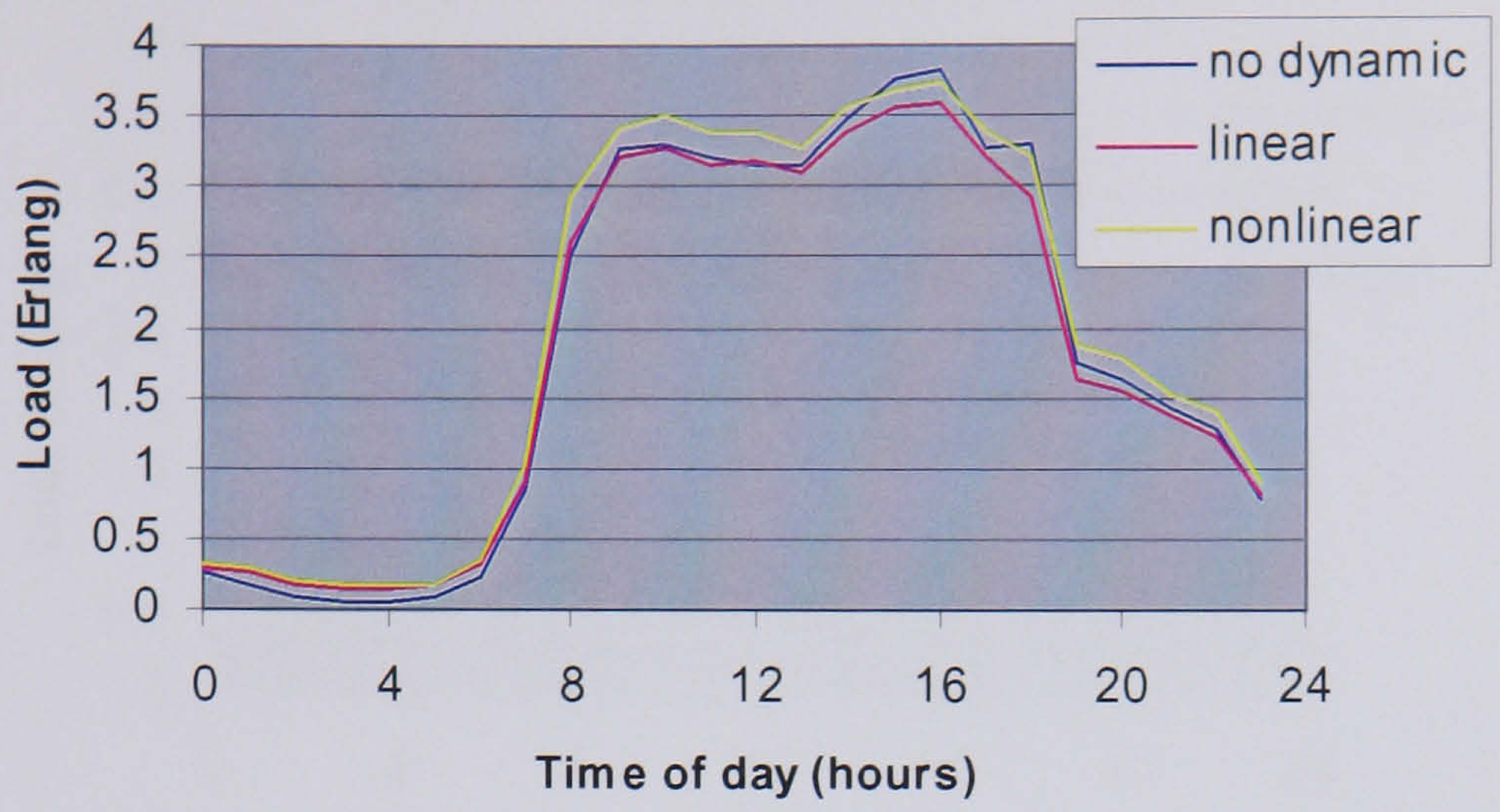

Figure 5-5 *Traffic distribution for* $\beta = 1.0$

As price elasticity β increases, however, the effect of the pricing function on the profile of the load becomes significant, particularly during off-peak hours. As can be expected, more calls are generated with dynamic pricing using either pricing function (see Figure 5-6 and Figure 5-7). The non-linear demand function offers a significant increase in the number of calls generated throughout the day, whereas the linear pricing function is more efficient at reducing the intensity of the traffic during peak hours.

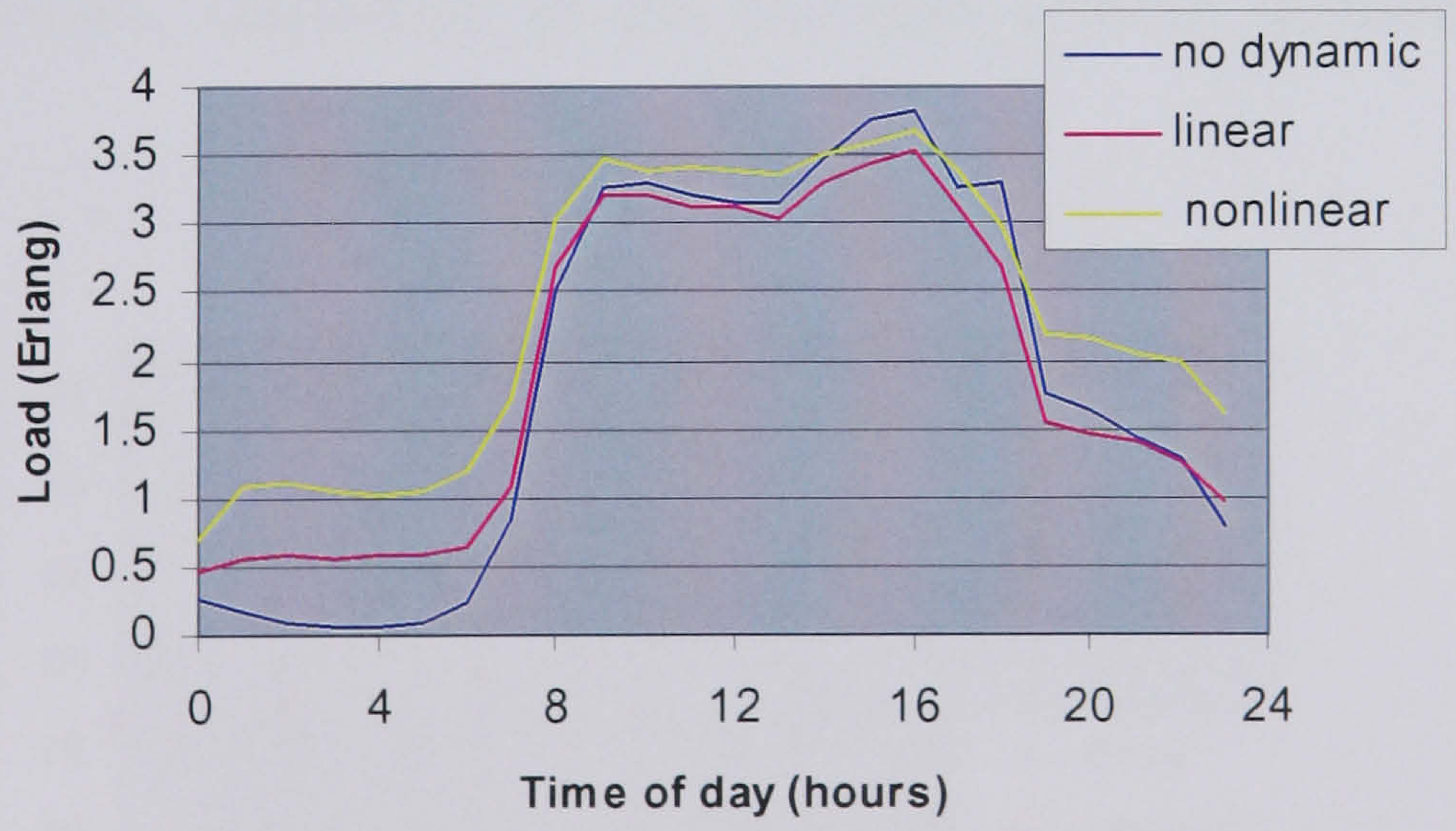

Figure 5-6 *Traffic distribution for* $\beta = 2.0$

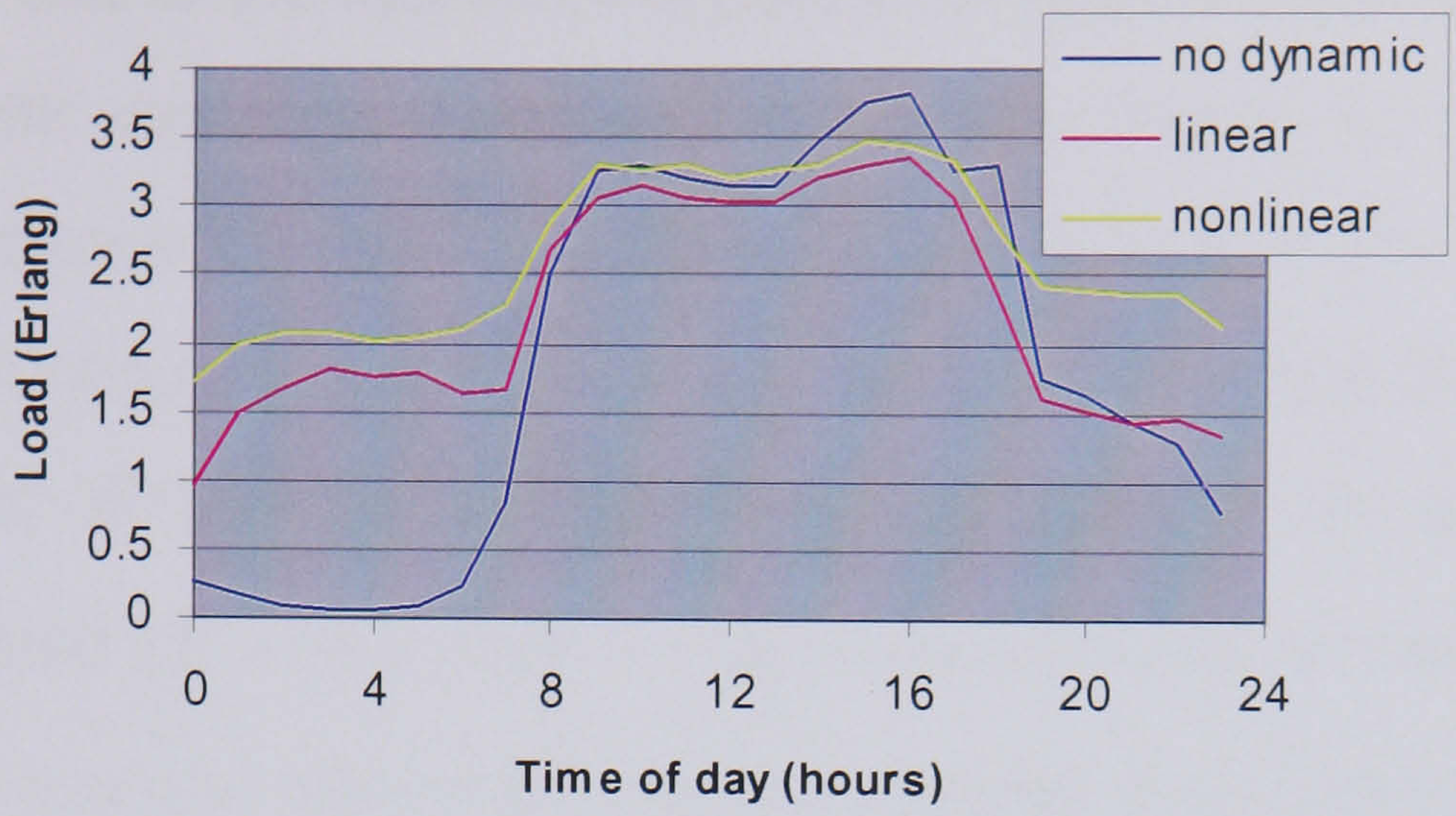

Figure 5-7 *Traffic distribution for* $\beta = 3.0$

5.3.1.2 User perspective.

The two pricing strategies also affect the welfare of the users, which is determined by the total number of successfully completed calls and the prices in the network.

Significantly more calls are serviced with the non-linear pricing function compared to the linear pricing function, for all values of $\beta > 0$. In fact, for $\beta \in [0.5, 1.5]$ the welfare of the users actually decreases with the linear demand function, compared to the scenario without dynamic pricing (see Figure 5-8).

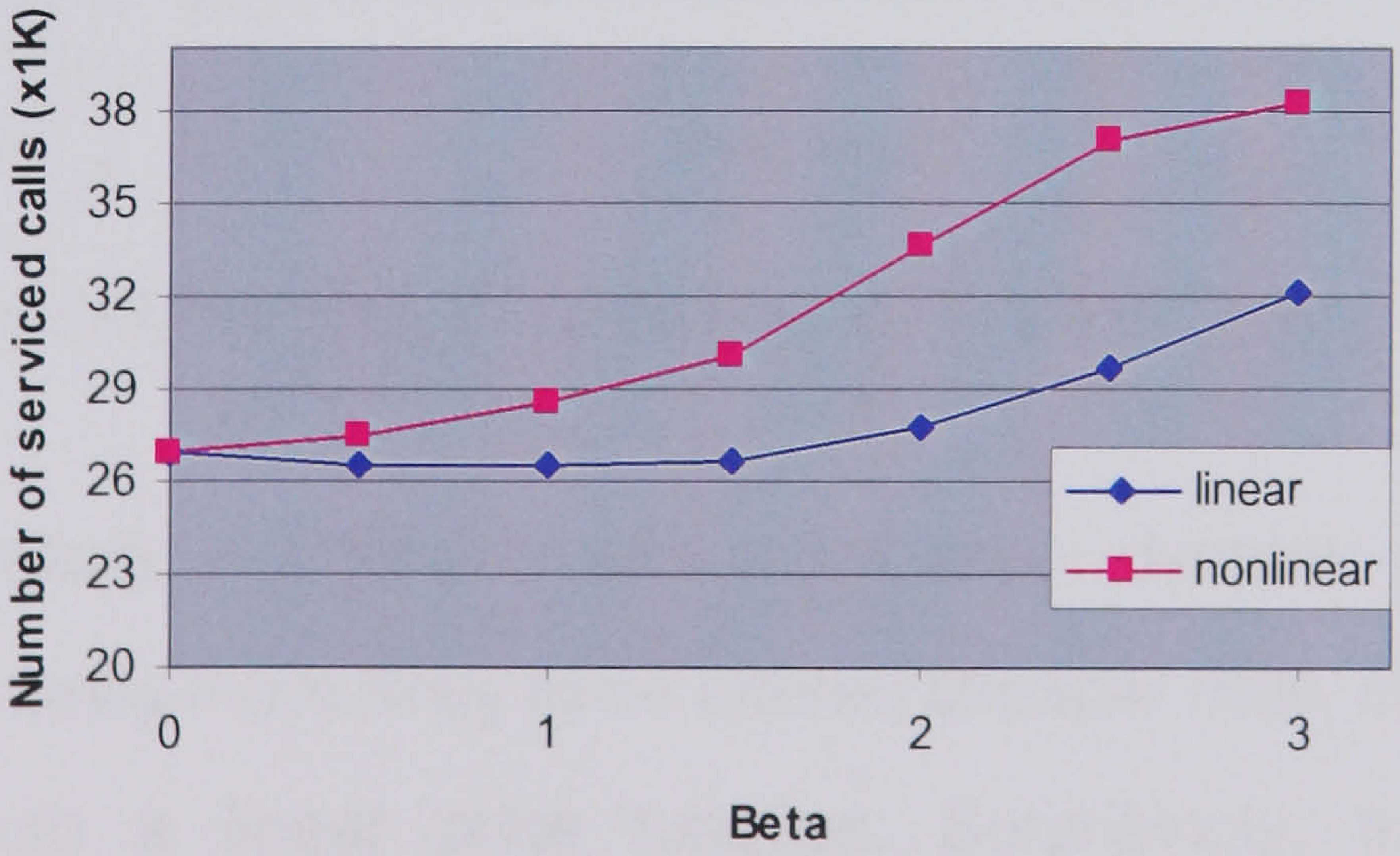

Figure 5-8 *Total number of successful calls as a function of* β

This is due to the fact that the additional traffic in off-peak hours is less than the traffic reduction during peak-hours (see Figure 5-5), leading to an overall reduction in the amount of traffic. However, for both the linear and non-linear price functions the number of successful calls significantly increases for $\beta \in [1.5, 3.0]$ by 18% and 41% respectively, compared to the situation without dynamic pricing ($\beta = 0.0$). This is due to the increase in user sensitivity to price, which leads to more calls being generated during off-peak hours than the calls suppressed during peak hours (see Figure 5-6 and Figure 5-7).

The distribution of most probable prices that users would see with the two pricing functions is also different, although a direct comparison is not feasible due to the different intermediate values. However, it can be noted that the probability of users seeing the lowest price (0.01 units) when the non-linear pricing function is operational is significantly higher than its frequency with the linear price function, for all β (see Figure 5-9 and Figure 5-10).

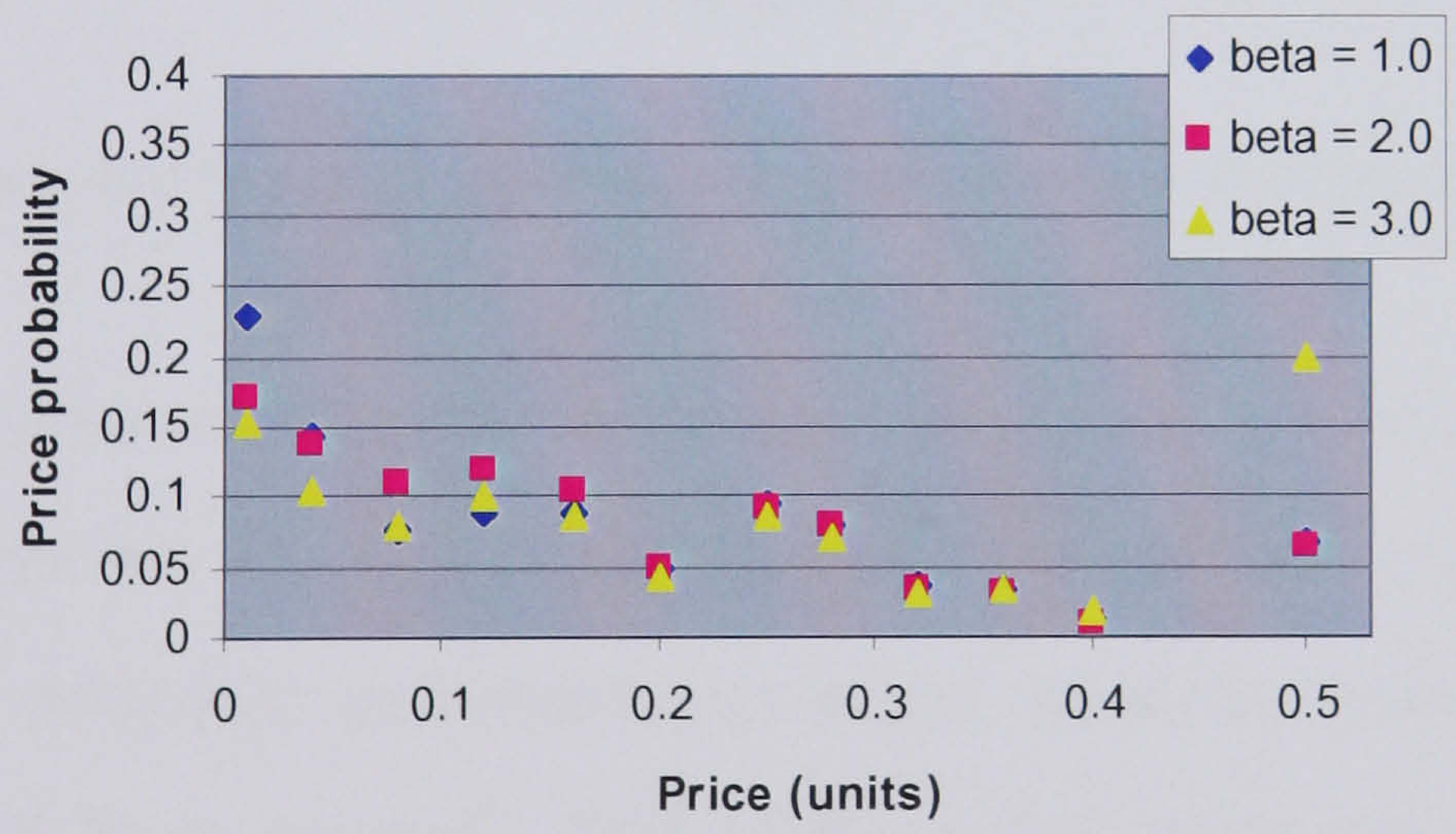

Figure 5-9 *Probability distribution of price for a linear pricing function*

Therefore, in a system with a non-linear dynamic pricing function users are significantly more likely to be offered cheaper calls, compared to users in a system with a linear price function. Surprisingly, the frequency of the appearance of the maximum price (0.5 units) is very similar with both the linear and non-linear pricing functions for inelastic and unit elastic demand ($\beta < 2.0$).

As β increases the probability distribution of the prices for the linear shifts towards the higher values of the prices, as an increase in user sensitivity leads to an increase in network load during off-peak hours. As demand becomes elastic ($\beta > 2.0$) there is a very significant increase in the probability of the maximum price (0.5 units), due to the significant increase in call blocking (Figure 5-4).

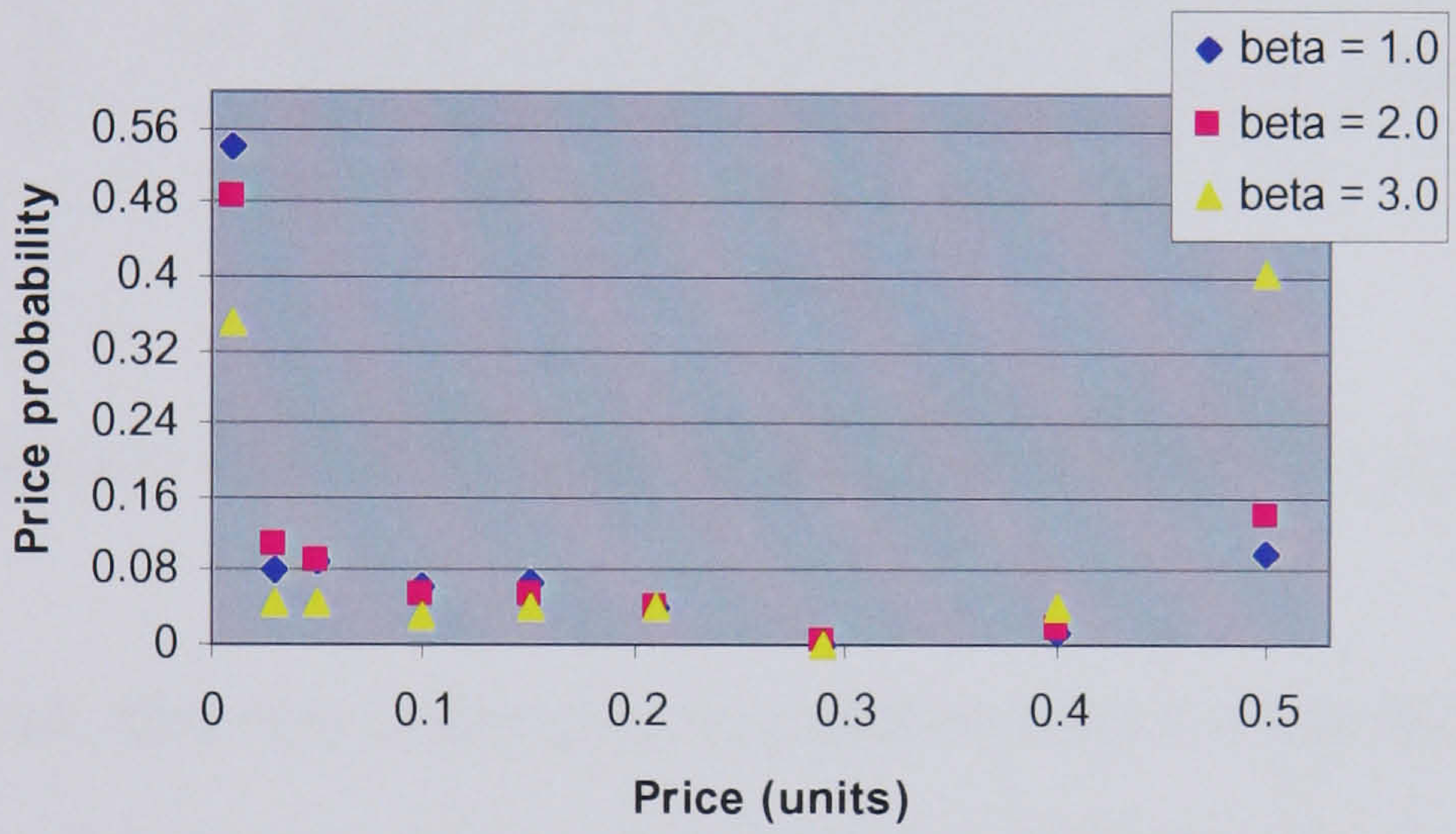

Figure 5-10 *Probability distribution of price for a non-linear pricing function*

The price distribution with the non-linear dynamic pricing function indicates that the users are most likely to see one of just two most probable prices (the minimum and maximum price), whereas intermediate prices are much less likely to appear[41]. This toggle effect arises because the non-linear pricing function does not offer sufficient control over call blocking and leads to an increase in the percentage of blocked calls. An increase of demand

[41] This situation is similar to the two price tariffs that users are accustomed to today in the UK, for example, and so this pricing would not appear particularly foreign to users.

elasticity β shifts the probability distribution of towards higher most probable prices, but this effect is only significant for $\beta > 1.5$.

The effect of demand elasticity β on the weighted average prices[42] is plotted in Figure 5-11.

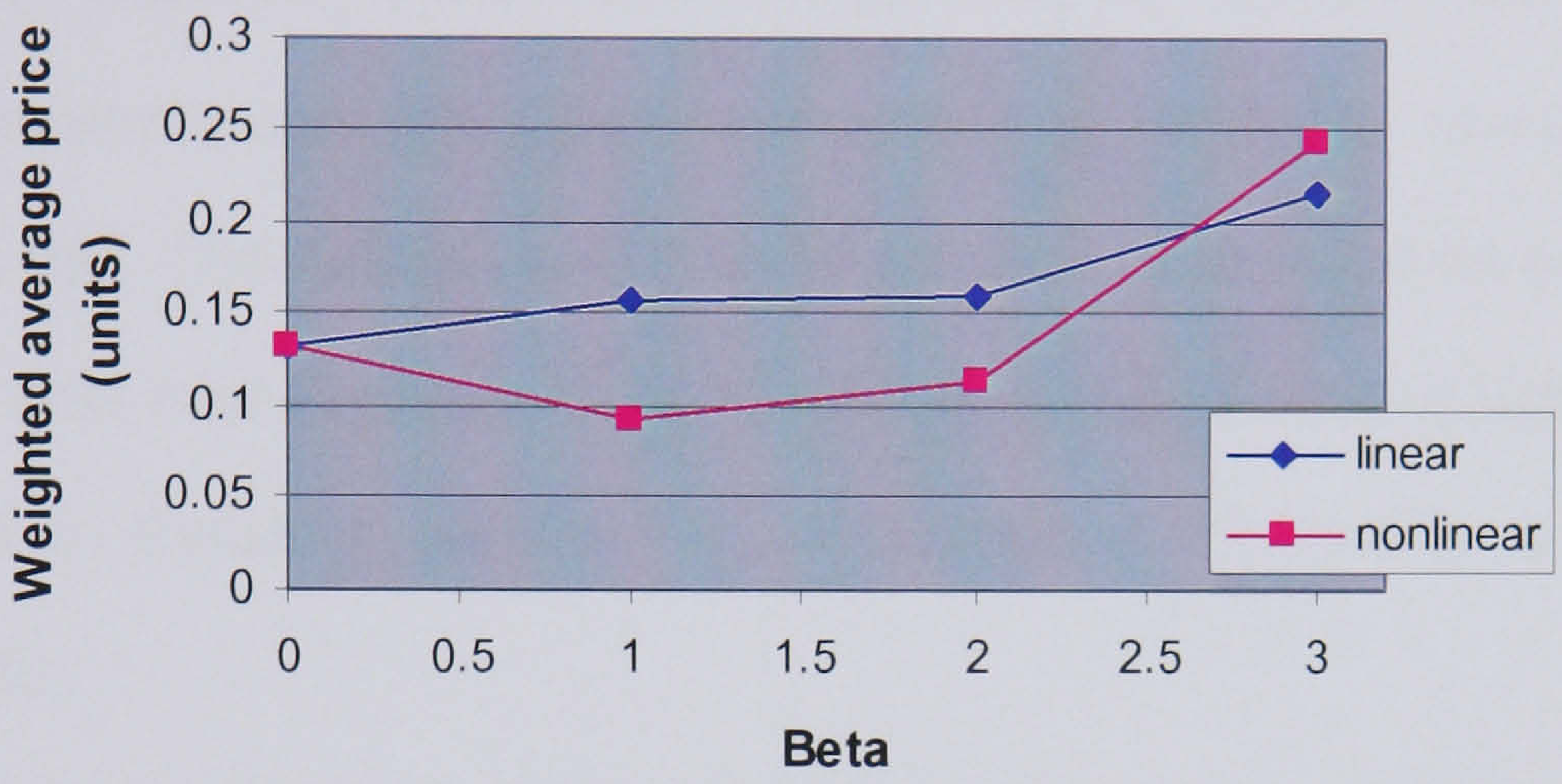

Figure 5-11 *Weighted average network prices as a function of* β

Overall, the non-linear pricing function offers a significantly lower average price than the linear pricing function for inelastic and unit elastic demand ($\beta \leq 2.0$) and would, therefore, be preferred by users. As demand becomes elastic ($\beta > 2.0$) the non-linear pricing function leads to higher average price than the linear function and, therefore, would become less desirable than the linear pricing function.

These results show that dynamic pricing can have a significant effect on the load in the system, the generated revenue and the percentage of blocked calls. Furthermore, the shape of the pricing function will affect the benefit users can derive from the system.

[42] The weighted average price is calculated by taking the probability of a price value occurring, multiplying it by the price value itself and adding all weighted price values together.

5.3.2 Dynamic Pricing and User Mobility.

An additional dimension to the problem specific to cellular networks is the mobility of users in reaction to the changing price, with users moving to cheaper cells and avoiding more expensive ones. This is modelled by altering the mobility elasticity parameter, defined as α in the mathematical model, which represents the willingness of users to move to cheaper cells before initiating a call. The results and graphs showing the reaction of the system with the linear and non-linear pricing functions at the introduction of user mobility are shown in detail in section F-2, Appendix F. Below is a summary of the main results.

The introduction of user mobility would lead to a decrease in the total revenue generated with both linear (Figure F-6) and non-linear (Figure F-7) pricing function as the elasticity of mobility α increases, due to the shift of a larger proportion of users to cheaper cells.

In addition, the percentage of blocked calls in the network can potentially increase (for $\alpha > 1.5$ and a linear pricing function (Figure F-8) and for $\alpha > 0.5$ with the non-linear pricing function (Figure F-9)). This is due to the increase in the proportion of users moving to cheaper cells as the mobility elasticity α increases. The higher proportion of users would saturate the available cell capacity and lead to an increase in the percentage of blocked calls.

Overall, the introduction of user mobility leads to a more even distribution of the network load as a function of time. However, fewer calls were being made with both linear and non-linear pricing functions. This is due to the reduction of calls in peak hours, as α increases, which is not compensated for by an increase in calls during off-peak hours. The only exception is the increase in the number of successful calls by 0.1% with a linear pricing function, $\alpha = 0.5$ and inelastic demand ($\beta = 1.0$) (Figure F-14).

User mobility has no effect on the average price in the network, which remains constant when a linear pricing function is used (Figure F-16). However, with the non-linear pricing function and inelastic demand ($\beta = 1.0$) the price steadily increases as α increases, raising by up to 8.9% for mobility elasticity $\alpha = 2.0$ (Figure F-17). For unit elastic demand ($\beta = 2.0$), the introduction of user mobility leads to an initial drop in the average price in the network by 3% (mobility elasticity $\alpha \leq 1.0$).

The results reported so far show that the response of the system to dynamic pricing is very complex and varied. The introduction of user mobility as a function of the changing cell prices adds an additional dimension to the problem, making accurate predictions about the effect of dynamic pricing even more difficult.

It should be noted, however, that in the simulations above, it was assumed that the parameters for mobility and price elasticity, α and β, are independent of each other and, therefore, their values were chosen independently. This, though, is not necessarily the case, and it is very likely that that the two coefficients are correlated. Although determining the exact relationship between the two coefficients is not within the scope of this project, the results suggest that the accurate estimation of these parameters is essential for reliable forecasting of the expected effect of dynamic pricing and, therefore, this area is recommended for future research

5.4 Chapter Summary.

This chapter suggested a market competition driven *ad hoc* dynamic price setting policy, taking into account minimum and maximum prices the network providers would want to charge as well as the value of the average price of competitors in the market. A seven-cell OPNET™ simulation model developed to test the effect of dynamic pricing was described and simulation results reported. The results showed that competition driven *ad hoc* dynamic pricing can be a very effective tool in increasing the revenue generated for the network operator (see Figure 5-3). The *ad hoc* pricing scenario also increases the total load in the network (see Figure 5-8) while reducing the number of blocked calls (see Figure 5-4). However, the shape of the pricing function will significantly affect the effectiveness of the dynamic pricing policy.

Additional factors, such as user mobility, were also shown to affect the effectiveness of dynamic pricing as a regulatory mechanism and these have to be considered before an accurate prediction of its efficiency can be made.

Chapter 6

The simulation results reported in the previous chapter show that dynamic pricing could be an effective tool for controlling network load and call blocking. However, its effectiveness depends on the shape of the pricing function. This chapter proposes a control theory approach to dynamic price setting. The most sensitive variables in the system are identified as the revenue generated by the network operator and the percentage of blocked calls. As a result these are recognised as being suitable controlled variables and a control system model for the cellular network suggested. Due to the non-linearity of the system traditional control theory methods are not suitable for finding the "best" pricing function applicable for the network. Therefore, the mathematical framework of a revenue attainment approach is then developed to solve the problem. The aim is to achieve a predefined target revenue for the network provider, initially assuming linearity of the dynamic pricing function. The linearity requirement is then dropped and a comprehensive model for the determination of optimal dynamic price setting strategy for any type of demand assumption is presented. Simulation results showing the effectiveness of the revenue attainment model are presented and discussed.

6.1 Control Theory Approach to Dynamic Price Setting.

Simulation results reported in Chapter 5 show that the suggested competition driven price setting methodology can lead to changes in system behaviour that are difficult to predict. For example, the linear pricing function will lead to an average price to users with inelastic demand (demand elasticity $\beta \leq 2.0$) of 0.15 to 0.16 units as β increases (see Figure 5-11), rather than the expected 0.22 units[43]. In this case the actual price is closer to the market average price of 0.13 units[44]. This pricing strategy also offers higher revenue for the network operators and at the same time reduces call dropping, without significantly affecting user welfare.

The non-linear pricing function on the other hand, leads to an average price of 0.09 to 0.11 units (as β increases), which is lower than the average market price. The lower price leads to an increase in the number of successful calls, but at the cost of significantly reduced revenue for the network operator and increased blocking in the network, which is undesirable from both network operator's and user's point of view.

As user demand becomes elastic, the weighted average price increases significantly for both functions (to 0.21 units and 0.24 units respectively). Although both are above the average market price of 0.13 units, they lead to an increase in both the number successful calls and revenue for the network provider, due to the very elastic demand. However, this is at the cost of a very significant (in the order of 600%, see Figure 5-4) increase in the percentage of blocked calls, which makes them unacceptable from a network operator's point of view.

These problems arise because the prices were chosen in an *ad hoc* manner, without taking into account any of the parameters affecting system behaviour such as the intensity and elasticity of user demand, expected call

[43] This price is derived using the pricing function in Figure 5-1 for $M = 1.0$

[44] Calculated using the Cellnet table (Occasional Caller +) in Appendix B.

length or load in the network. A more rigorous approach, taking into account these parameters, is necessary to ensure that the prices set will lead to the desired system behaviour.

Control theory provides a powerful tool for determining the control necessary to steer a system into its desired state. To define the cellular network as a control system we have to find the most suitable controlled variable or variables z and identify any significant disturbances w.

The results presented in section 5.3 highlight how sensitive a network provider's revenue and the percentage of blocked calls are to the choice of pricing function. If, for example, the network operator implements the non-linear pricing function, it would lead to a significant drop in their revenue. Decreased revenue will directly affect the profit and investment of the network operator and in the long run will have a negative effect on the network operator's market share as pointed out in section 3.1.1. The percentage of blocked calls would, on the other hand, affect users' perception of the QoS provided by the network operator. Since the generated revenue and the percentage of blocked calls are particularly sensitive variables, this also means they could be good candidates to be controlled variables in the system.

Another parameter that significantly affects the network behaviour once dynamic pricing is implemented is the magnitude of the load in the network as a function of the time of day, as this will have a bearing on the performance of the pricing function. For example, the effectiveness of the linear *ad hoc* function for revenue generation and control of call blocking decreases as the magnitude of the peak load in the network decreases. The system was tested with the linear *ad hoc* pricing function defined in Figure 5-1 and a reduced intensity of the traffic during peak hours by 50 (for medium load) and by 100 (for small load) calls per hour (see Figure 6-1).

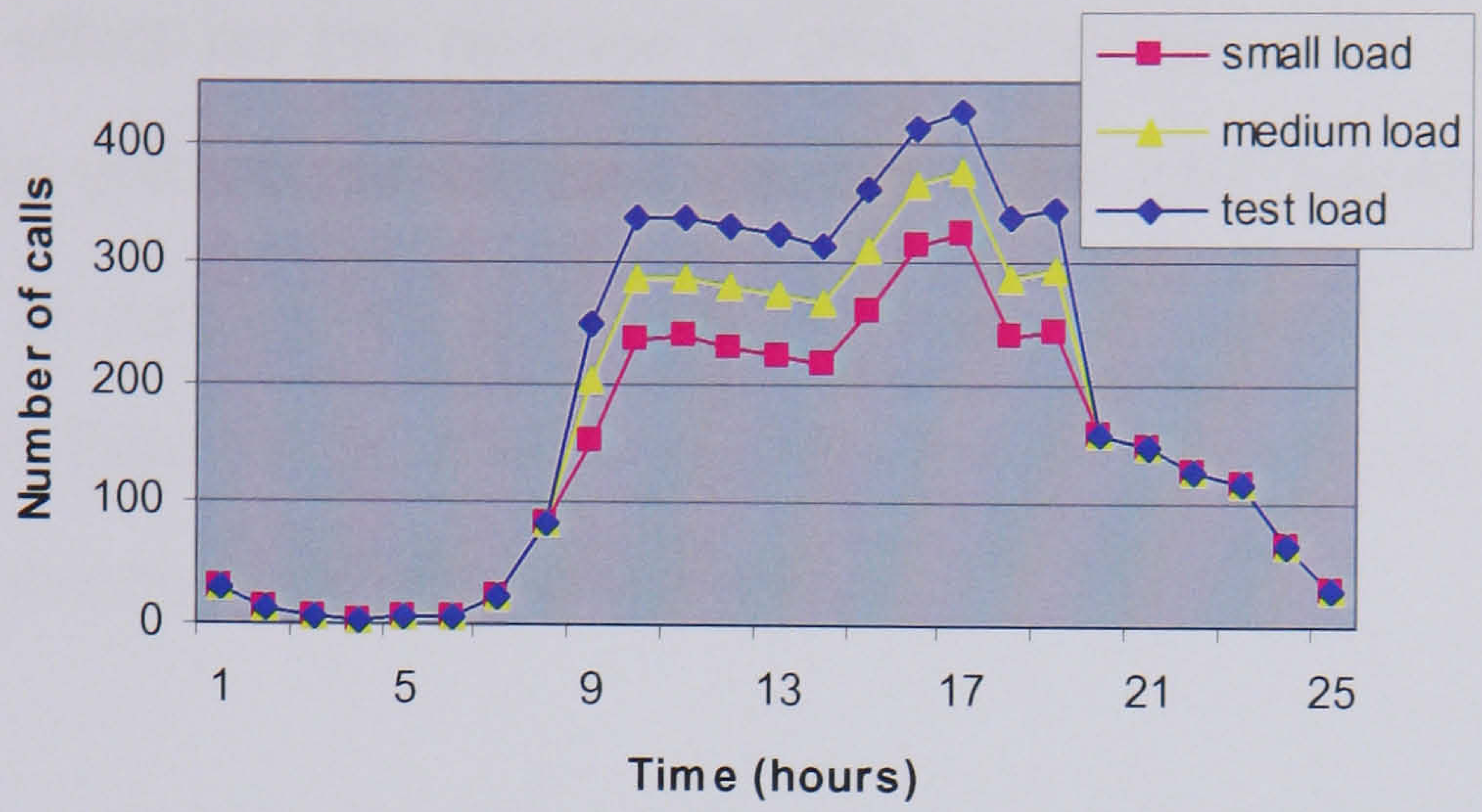

Figure 6-1 *Comparison of different peak-hour traffic loads.*

With a small network load the percentage of blocked calls, in fact, increased with the introduction of dynamic pricing (see Figure 6-2) for all values of the demand elasticity β. This is due to the over-stimulation of peak-hour demand, which the linear pricing function cannot control efficiently at very low magnitudes of the peak-hour load in the network.

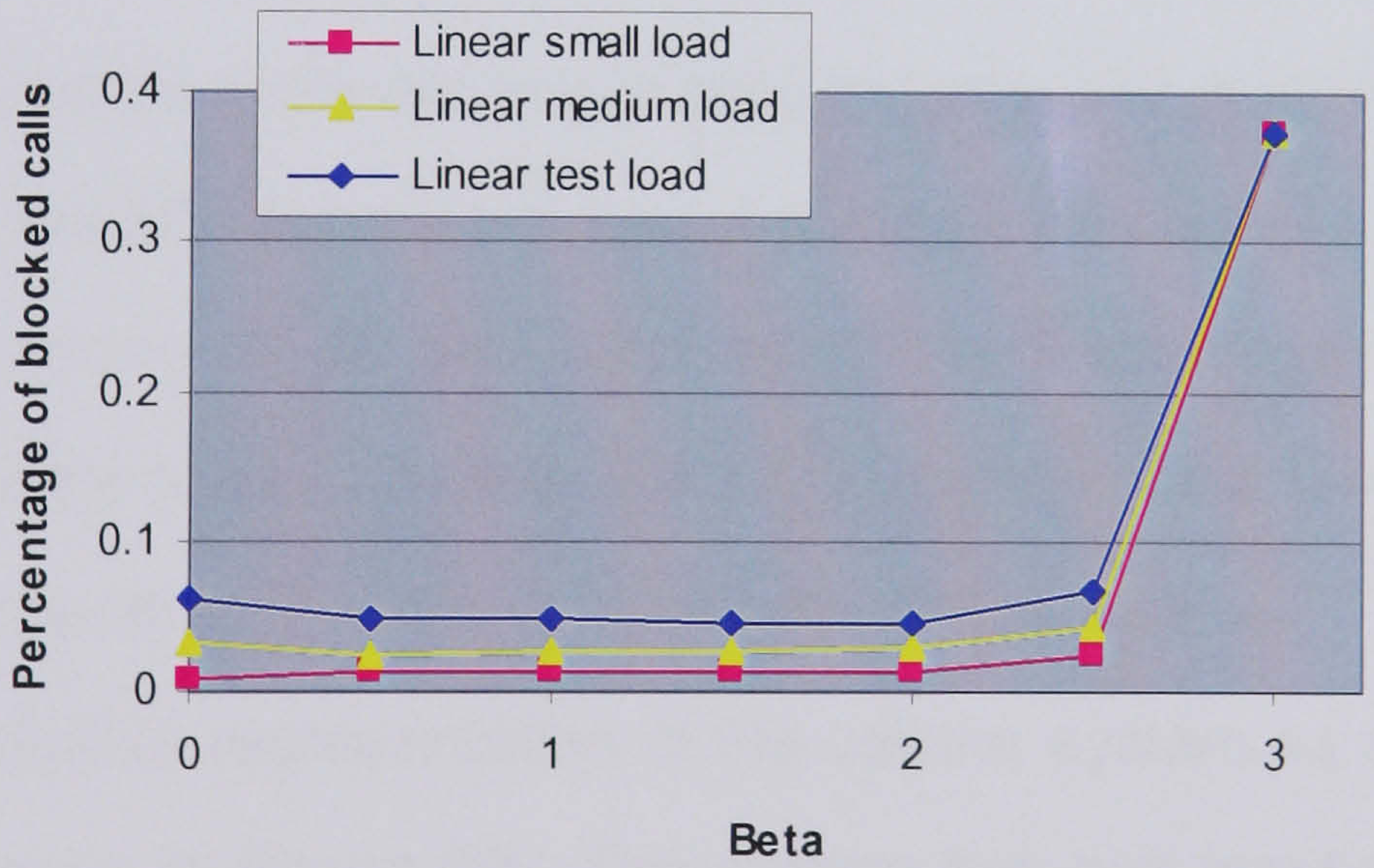

Figure 6-2 *Comparison of call blocking with different peak-hour traffic loads*

However, as the magnitude if the peak-hour load increases the effectiveness of the linear pricing function in controlling the system improves and the percentage of blocked calls decreases.

The effect on the revenue is also significant. With a small load and inelastic or unit elastic demand ($\beta \leq 2.0$), the total generated revenue will decrease or remain the same as the revenue generated without dynamic pricing (see Figure 6-3). With a larger load, however, (medium or test load) the expected revenue will actually increase.

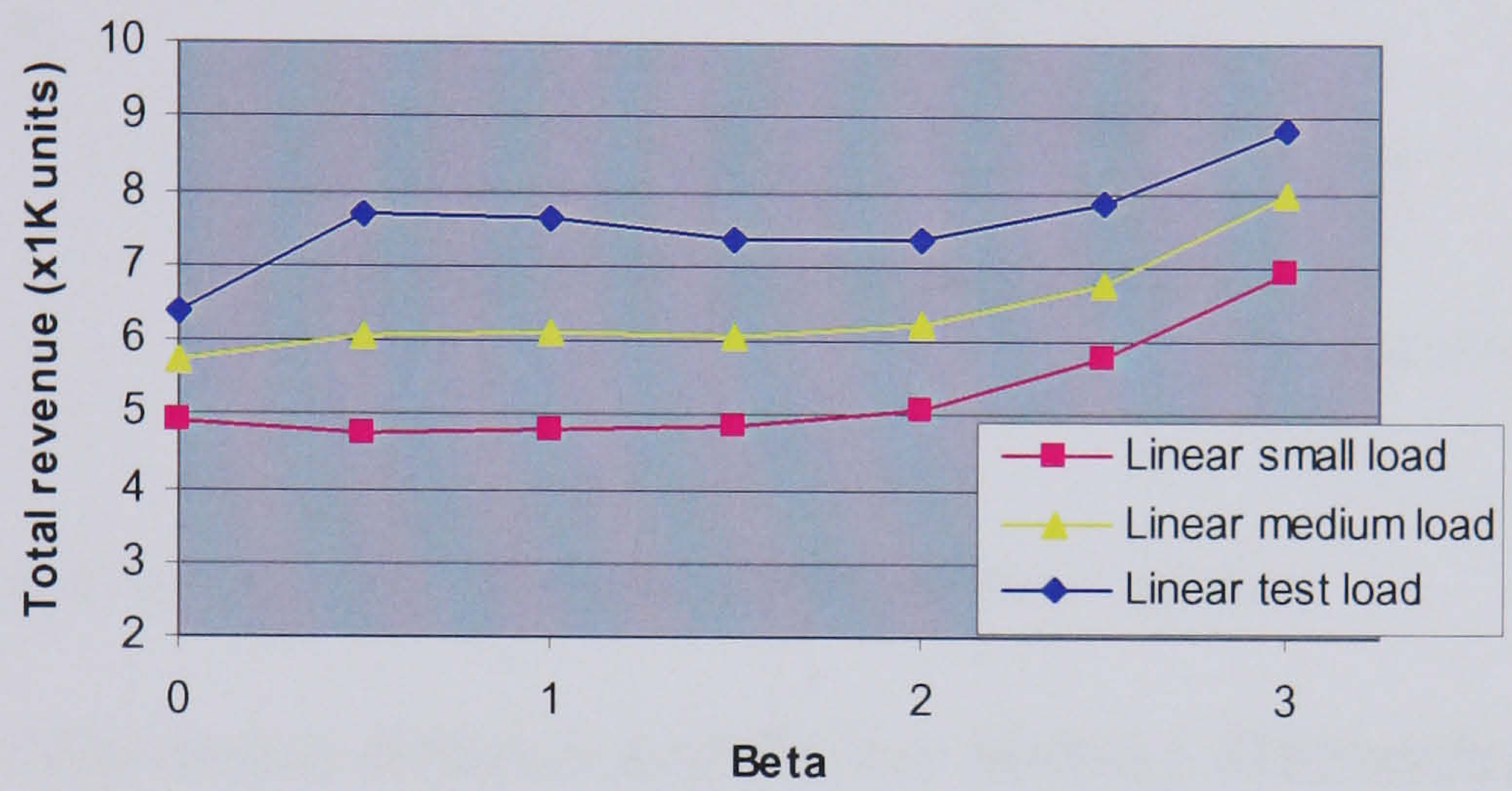

Figure 6-3 *Comparison of generated revenue with different peak-hour traffic loads*

These results show the importance the intensity of the traffic load plays in determining the effectiveness dynamic pricing as a control function. Therefore, to derive a pricing model that will lead to the desired system behaviour, the magnitude of the load in the network as a function of the time of day has to be taken into account in the form of a system disturbance w.

One possible representation of the cellular system as a feedback control system is shown in Figure 6-4. The system has one input (demand) and two outputs: revenue and network load, which act as controlled variables. This is a non-linear control system which cannot be solved using traditional control theory methods because the principle of superposition (see section 2.3) cannot be applied (Phillips and Harbor [104]). Therefore, alternative mathematical methods have to be considered.

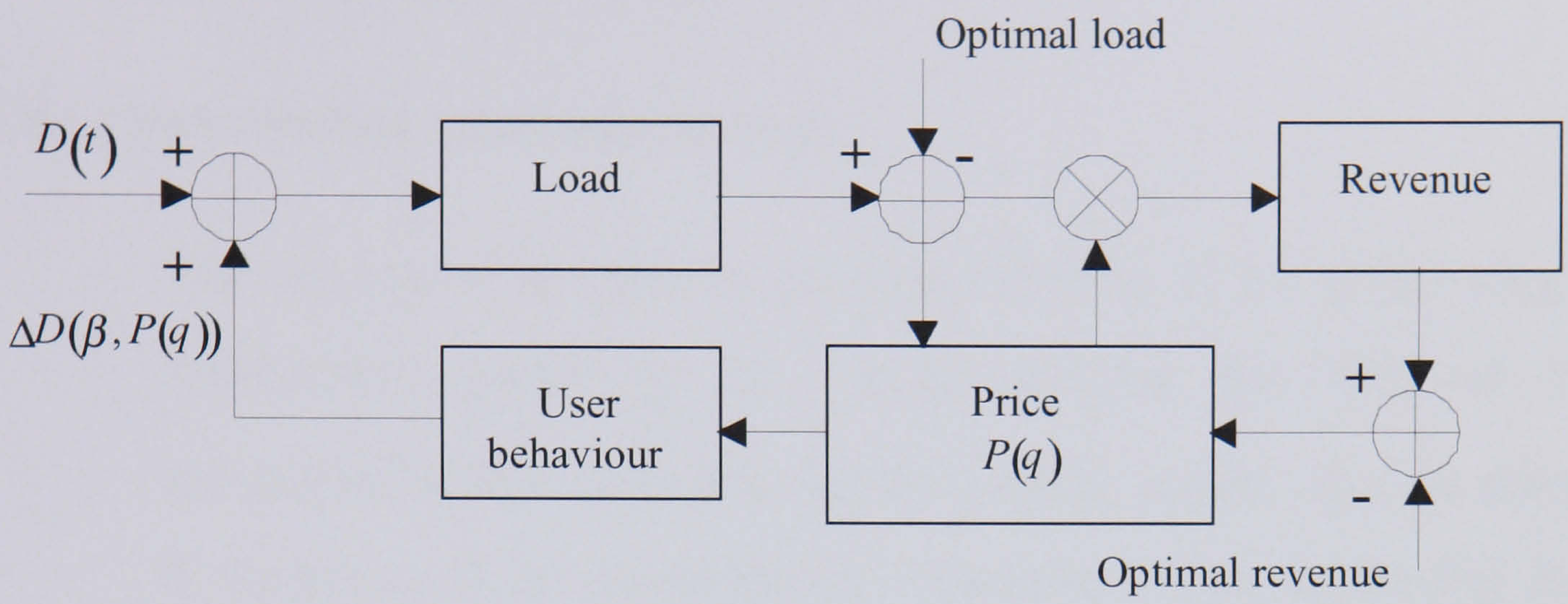

Figure 6-4 *The cellular network as a control feedback system.*

6.2 Linear Revenue Attainment Price Setting Methodology.

6.2.1 Revenue Attainment Mathematical Model Definition.

The objective is to define and solve a mathematical model describing the system depicted in Figure 6-4 to establish a pricing function the network operator has to use. The pricing function has to ensure that certain target revenue[45] $R_{desired}$ is generated within a fixed time interval (White [105]), while the load in the network is optimal.

In mathematical terms the revenue generation objective can be expressed as:

$$\text{Minimise } R_{desired} - \int_{t_0}^{t_1} R_{current}(t)dt \qquad (6\text{-}1)$$

$R_{desired}$ - Desired revenue;

[45] Network providers can determine the required revenue by, for example, using equation (4-28).

$R_{current}(t)$ - Revenue generated at time t;

t_0 - Start of optimisation period;

t_1 - End of optimisation period.

6.2.1.1 Mathematical model assumptions

1. This approach can lead to potential increase of the prices over the optimisation period $(t_1 - t_0)$. Target revenue that has not been generated at the beginning of the pricing period, due to demand fluctuation, will be generated by increasing prices at the end of the day. This could be perceived as "unfair" by both the industry watchdog and the users who will expect cheaper prices at the end of the day. To avoid the problems of uncertainty, the network operator can make the increments in revenue constant. This would prevent potential cross subsidy of prices over time. In addition, the value of the desired revenue would be updated on hourly basis.
2. For the purpose of this study it will be assumed that the network operator is attempting to match the revenue generated without dynamic pricing to the revenue generated with dynamic pricing. This choice will minimise any competitive disadvantage the network operators can incur at the introduction of dynamic pricing, as their consumers will be spending overall the same on calls as before dynamic pricing was introduced[46]. The target revenue generated using the OPNET™ simulation without dynamic pricing is plotted in Figure 6-5.

[46] This assumption has to be checked more rigorously before the practical application of dynamic pricing.

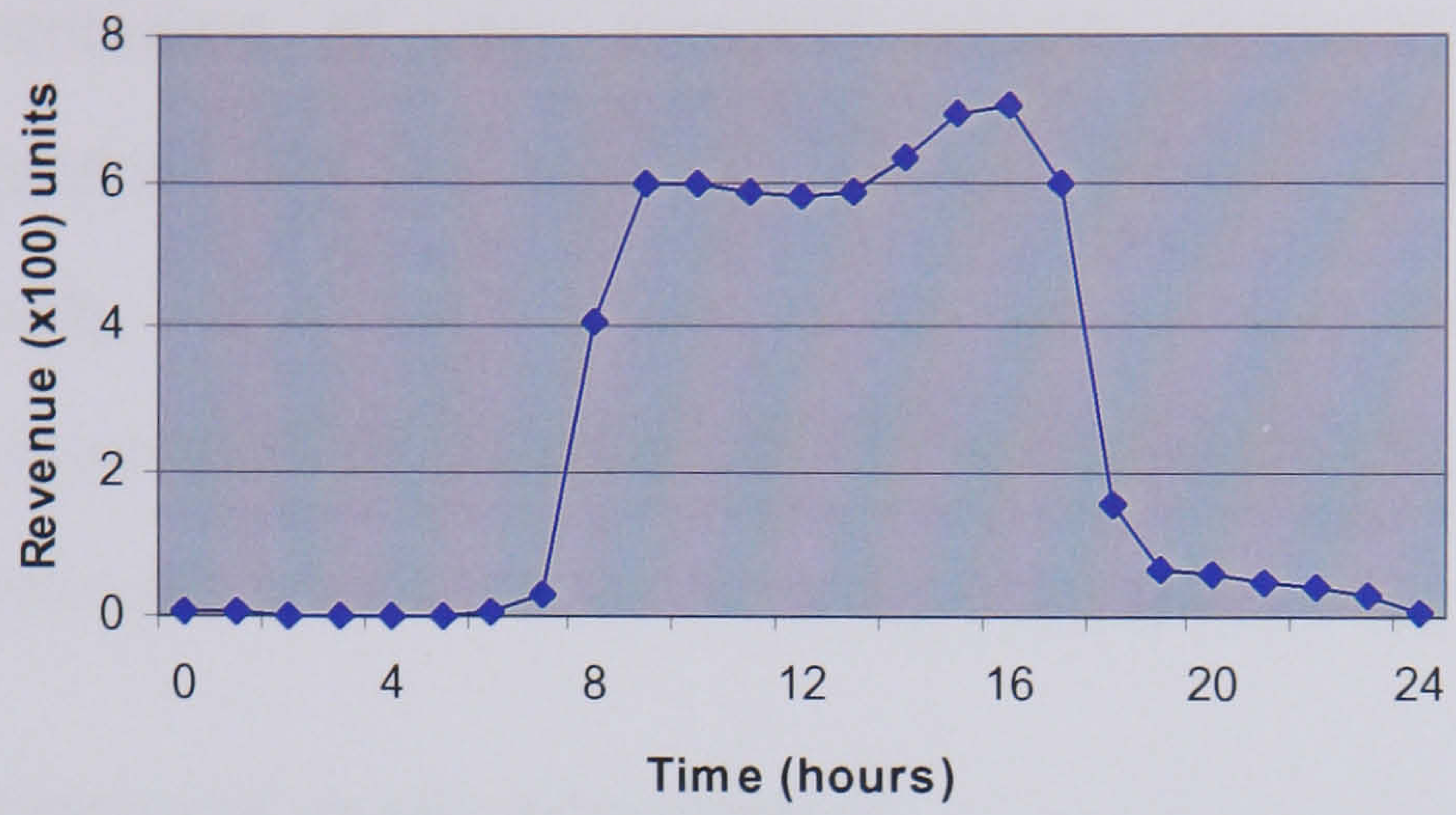

Figure 6-5 *Hourly revenue attained without dynamic pricing.*

3. The pricing function will be assumed to be linear (equation (6-2)).
4. The price in each cell will depend on the instantaneous load in the cell q and will be fully described by the gradient ψ of the dynamic pricing function.

$$P(q) = \psi q + P_{\min} \qquad (6\text{-}2)$$

q - Instantaneous load in the cell;

ψ - Price gradient;

$P_{\min}$ - Minimum price charged, assumed to be 0.01 price units.

5 It will be assumed that the proportion of the load contributed by each individual cell in the network is equal. Therefore, as the instantaneous load in the network is the sum the instantaneous loads in the individual cells, the total load in the network will be proportional to the total load in the individual cells. In this case, the two terms are interchangeable. In practice, this assumption will not be accurate for the entire network and the network operator will have to calculate the revenue optimisation prices on per cluster basis, for example, to reduce the expected price estimation error.

6 Due to the limited amount of data available regarding the temporal distribution of user demand without dynamic pricing, it will be assumed that the expected number of users in the network changes hourly. As a result of this assumption, the demand function will be independent of the state of the intermediate price update interval t within the hour and can be written as $D(\beta,(\psi q + P_{\min}))$.

6.2.1.2 Mathematical model development.

Given a desired hourly revenue of $R_{desired}$ and T price update subintervals, the problem can be defined as:

$$\text{Minimise } \int_{t_0}^{t_1}\left(\frac{R_{desired}}{T} - R_{current}(t)\right)^2 dt \qquad (6\text{-}3)$$

T - Number of price update subintervals;

t - Intermediate price update interval;

$R_{current}(t)$ - Total revenue generated at time t;

t_0 - Start of optimisation interval;

t_1 - End of optimisation interval.

Taking the square of the error between the two terms in the equation ensures that both the positive and negative error is minimised.

The revenue generated at each pricing subinterval T with the linear pricing function (6-2) is:

$$R_{current}(t) = \frac{(D(\beta,(\psi q + P_{\min})))}{T}(\psi q + P_{\min})\tau \qquad (6\text{-}4)$$

$D(\beta,(\psi q + P_{\min}) - P_{bias})$ - Expected compound demand for network services;

β - Elasticity of demand;

τ - Average call-holding time;

q - Instantaneous load in the network;

ψ - Price gradient;

$P_{\min}$ - Minimum price charged, assumed to be 0.01 price units..

In addition to the requirement for revenue generation, the network provider has to ensure that the capacity of the network is not exceeded by demand. This condition ensures that the derived prices encourage optimal resource usage without call blocking. In mathematical terms this is expressed as a constraint ensuring that the number of calls at any given time t is less or equal to the available network capacity C:

$$\frac{D(t,\beta,(\psi q + P_{\min}))}{T} - \frac{q}{\tau} \leq C \qquad (6\text{-}5)$$

C - Network capacity as number of channels.

In fact, from network operator's point of view, equality is particularly desirable, as it will encourage more network usage, and potentially more revenue when the available resources are under-utilised. Therefore, to find the revenue attainment price P_R equations (6-3) and (6-5) have to be solved simultaneously using the Lagrangian multiplier. This requires the optimisation of:

$$L = \left(\frac{R_{desired}}{T} - \frac{(D(\psi,q))}{T}(\psi q + P_{\min})\tau \right)^2 + \lambda \left(C - \frac{D(\psi,q)}{T} + \frac{q}{\tau} \right) \qquad (6\text{-}6)$$

Taking the derivative with respect to the variables ψ, q and λ, equating the resulting equations to 0, and removing λ results in:

$$-2\tau\left(\frac{R_{desired}}{T} - \frac{(D(\psi,q))}{T}(\psi q + P_{\min})\tau \right)$$
$$\left(\frac{\psi D(\psi,q)}{T} + \frac{(q\psi + P_{\min})}{\tau} + D(\psi,q)q\left(\frac{1}{\tau}\frac{d\psi}{dD(\psi,q)} - \frac{d\psi}{dq}\frac{1}{T} \right) \right) = 0 \qquad (6\text{-}7)$$

$$C - \frac{D(\psi,q)}{T} + \frac{q}{\tau} = 0 \qquad (6\text{-}8)$$

Equation (6-7) has two independent solutions and can be split into two independent systems of equations, (6-9) and (6-10) respectively:

$$\left| \begin{array}{l} \frac{R_{desired}}{T} - \frac{D(\psi,q)}{T}(\psi q + P_{\min})\tau = 0 \\ C - \frac{D(\psi,q)}{T} + \frac{q}{\tau} = 0 \end{array} \right. \qquad (6\text{-}9)$$

$$\left| \begin{array}{l} \frac{\psi D(\psi,q)}{T} + \frac{(q\psi + P_{\min})}{\tau} + D(\psi,q)q\left(\frac{1}{\tau}\frac{d\psi}{dD(\psi,q)} - \frac{d\psi}{dq}\frac{1}{T}\right) = 0 \\ C - \frac{D(\psi,q)}{T} + \frac{q}{\tau} = 0 \end{array} \right. \qquad (6\text{-}10)$$

Equation (6-9) represents a simplified version of the original problem and its solutions will involve $R_{desired}$, the capacity of the network and all necessary variables. It solution will be discussed in the following section.

Equation (6-10), on the other hand, does not explicitly contain $R_{desired}$ and, in addition, solutions to this equation give an unexpected result with the price gradient $\psi < 0$. This can be verified by expressing the load q from the second equation in (6-10), calculating $\frac{dq}{d\psi}$ and substituting in the first equation of the system (6-10) which gives:

$$\psi = -\frac{P_{\min}T}{D(\psi,q)\tau + qT} \qquad (6\text{-}11)$$

All variables in (6-11) are positive and, therefore, the price gradient ψ is negative. Thus the solution of equation (6-10) will be an optimal pricing function with a negative slope, *i.e.* the price of the calls will fall as the load in the network increases. This is certainly highly undesirable from a network operator's point of view because it will lead to even busier network in peak hours. Further research is necessary to establish the reasons for this result and so system (6-10) will not be used to determine the optimal price gradient ψ in this thesis.

6.2.2 Revenue Attainment Model Solution.

The equations presented in (6-9) cannot be solved analytically because of their non-linear nature. However, they can be solved numerically by introducing a penalty term in the equation. Penalty functions are mathematical tools which impose an increased penalty for violating a constraint (Nash and Sofer [106] or Gill *et al.* [107]). The revenue attainment problem can be solved by choosing one of the equations in (6-9) as the function to be minimised, while the other equation is imposed as a constraint. To ensure that the constraint is met a penalty function, such as the quadratic loss penalty function $\frac{1}{2}\rho(g(x))^2$, is introduced and the problem is solved for increasing values of the penalty parameter ρ. Applying this to (6-9) leads to the minimisation problem:

$$\text{Min}\left(\frac{R_{desired}}{T}-\frac{D(\psi,q)}{T}(\psi q+P_{\min})\tau\right)+\frac{1}{2}\rho\left(C-\frac{D(\psi,q)}{T}+\frac{q}{\tau}\right)^2 \qquad (6\text{-}12)$$

The choice of the main equation and the constraint is not binding and the minimisation function could have equally been defined by swapping the positions of the function and the constraint. Both Matlab and C programs for finding the optimal price gradient ψ are suggested in Appendix D.

Using the Matlab program and the revenue data in Figure 6-5 it was established that the revenue attainment prices very rarely satisfied the equality of the capacity constraint, which also maximises network utilisation. This indicates that an exact solution to this optimisation problem using the attained revenue from the simulations is not always possible. Defining a price which will attain a given revenue leads to a price which will not offer maximum capacity optimisation over all hours in the day and vice versa. This is due to the fact that the number of users affected by the dynamic prices is not constant but changes as a function of time of day. During off-peak hours the proportional increase in demand due to the fall in price will not lead to the same proportional increase in the number of calls made by users, and this will lead to a reduction in the generated revenue[47]. As the simulations were based on real life data, this suggests that a network provider would also be faced with this problem.

In such a case, network providers would have to look for a compromise between revenue attainment and maximum network capacity utilisation and determine the actual price in the network based on their preference for revenue or capacity utilisation. One possible resolution is the use of a combination of the optimal revenue attainment price $p_R(q)$ and the optimal capacity utilisation price $p_C(q)$ for a given network load q. These prices can be linked by the addition of parameters representing the network operator's preference for either revenue attainment (ε, say) or capacity utilisation (ϕ, say) and the results normalised.

$$p_{optimal}(q)=\frac{\varepsilon p_R(q)+\phi p_C(q)}{\varepsilon+\phi} \qquad (6\text{-}13)$$

It was also observed that the price gradient ψ and the instantaneous load q are inversely linearly dependent and, therefore, in order to determine the gradient ψ, the load q has to be known. Using the demand for network services without dynamic pricing, a network operator could predict the expected load in the network (q_{fixed}, say). The network operator would then be

[47] In this thesis it was assumed that users have a unit inelastic demand and, therefore, small changes in the price lead to small changes in demand. Whether this assumption is totally justified will require further research.

able calculate the optimal price $P_{optimal}(q_{fixed})$ for this level of demand, using (6-12). From this the gradient of the optimal pricing function can be derived using:

$$\psi = \frac{P_{optimal}(q_{fixed}) - P_{\min}}{q_{fixed}} \qquad (6\text{-}14)$$

$P_{optimal}(q_{fixed})$ - Optimal price at network load q_{fixed}.

Substituting (6-13) in (6-14) the optimal gradient ψ of the pricing function for a particular expected network load q_{fixed} can be found. This process can be seen as a calibration of the pricing for a particular load in the network.

Once the gradient of the pricing function is determined the complete pricing function for all states of the network can be derived:

$$P(q_{current}) = \left(\frac{\left(\frac{\varepsilon P_R(q_{fixed}) + \phi P_C(q_{fixed})}{\varepsilon + \phi} \right) - P_{\min}}{q_{fixed}} \right) q_{current} + P_{\min} \qquad (6\text{-}15)$$

ε - Network operator's preference for revenue attainment;

ϕ - Network operator's preference for capacity optimisation;

q_{fixed} - Expected load for optimal price calibration;

$q_{current}$ - Current load in the network;

$P_R(q_{fixed})$ - Revenue attainment optimal price at load q_{fixed};

$R_C(q_{fixed})$- Capacity attainment optimal price at load q_{fixed};

$P_{\min}$ - Minimum price charged, assumed to be 0.01 price units.

Equation (6-15) derives the optimal price for an instantaneous network load $q_{current}$, given a network operator's preferences for revenue attainment ε and maximum capacity utilisation ϕ. This pricing strategy was implemented in the OPNET™ model developed in section 5.2, which was used to test its effectiveness. The simulation results are reported in the next section.

6.2.3 Simulation Results with Linear Revenue Attainment Pricing.

This section gives results from the OPNET™ simulation, which was run for low, medium and high network operator preference for revenue attainment ε (parameter $\varepsilon = 0.1$, 0.5 and 0.9 respectively). Since, by definition, the parameter for capacity utilisation $\phi = 1 - \varepsilon$, the simulation will test corresponding high, medium and low operator preference for capacity utilisation (parameter $\phi = 0.9$, 0.5 and 0.1 respectively). The effectiveness of the revenue attainment pricing algorithm in controlling the system would be tested by observing the behaviour of the controlled variables: revenue, load and percentage of blocked calls

In Figure 6-6 the total revenue generated in the network at different values of the demand elasticity β and network operator's revenue attainment preference ε is shown. Overall the total desired revenue (in this case the revenue achieved without dynamic pricing *i.e.* demand price elasticity $\beta = 0.0$) is never achieved, with the overall revenue on average 25% below the target revenue. Surprisingly, the best result is achieved in the case of low preference for revenue attainment (parameter $\varepsilon = 0.1$) and unit elastic demand (demand elasticity $\beta = 2.0$). In fact, for unit elastic demand, the generated revenue actually decreases (by 3.6%) as the network operator's preference for revenue attainment increases ($\varepsilon = 0.9$).

Although this decrease is not statistically significant, it suggests that the effect of the maximum capacity utilisation price on the overall dynamic price becomes more significant as the elasticity of demand β increases. This is confirmed by the fact that for inelastic demand ($\beta = 1.0$) the total revenue increases, as expected, when the revenue attainment preference ε increases (by 2.6% for $\varepsilon = 0.5$ and 1.5% for $\varepsilon = 0.9$).

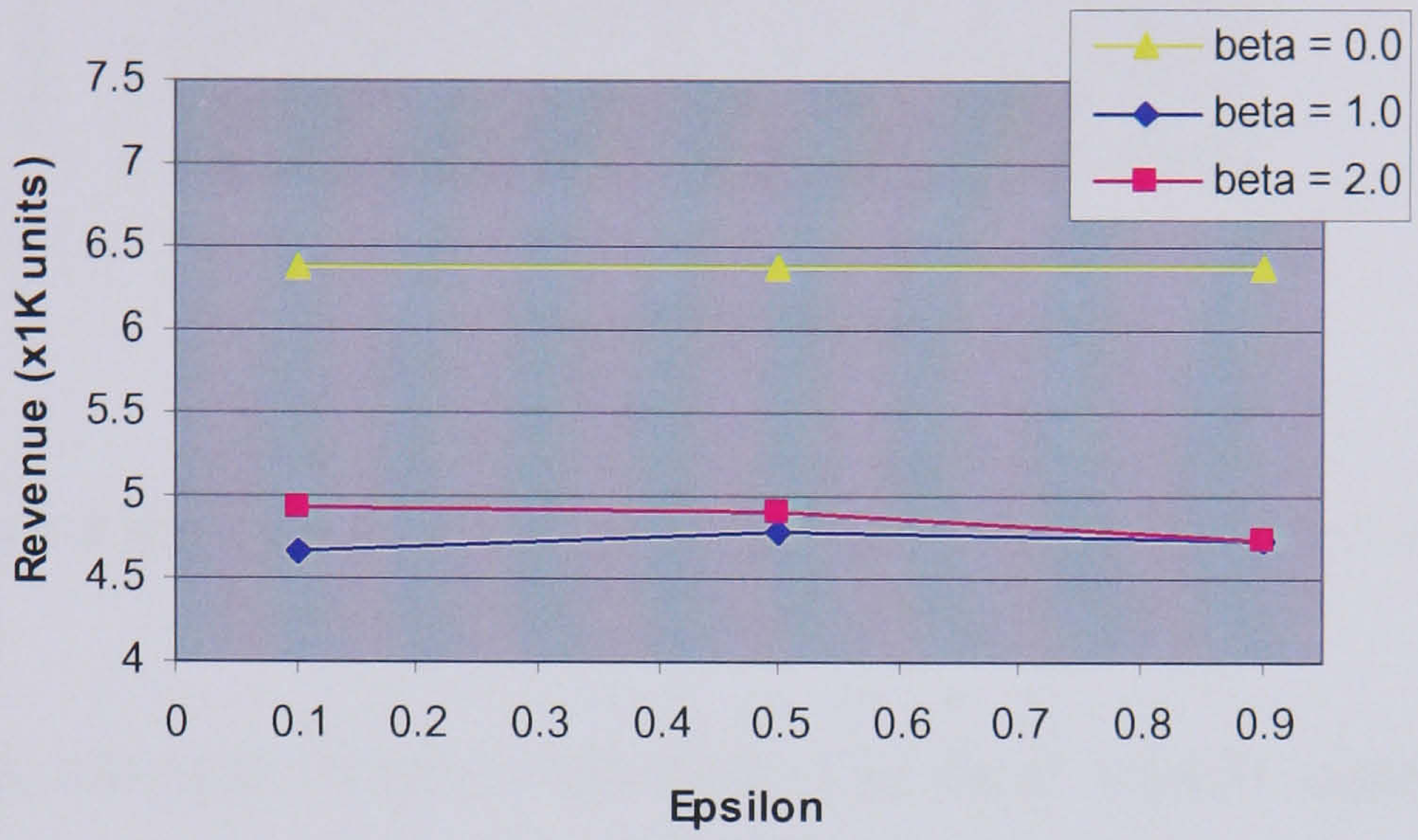

Figure 6-6 *Total revenue as a function of the revenue attainment preference* ε

The revenue generated as a function of the time of the day is compared to the desired revenue in Figure 6-7 for inelastic (price elasticity $\beta = 1.0$) and Figure 6-8 for unit elastic (price elasticity $\beta = 2.0$) demand. The revenue attainment strategy does not generate the desired revenue even with the highest preference for revenue attainment ε, with discrepancy between target and achieved revenue being most significant in both cases during peak hours.

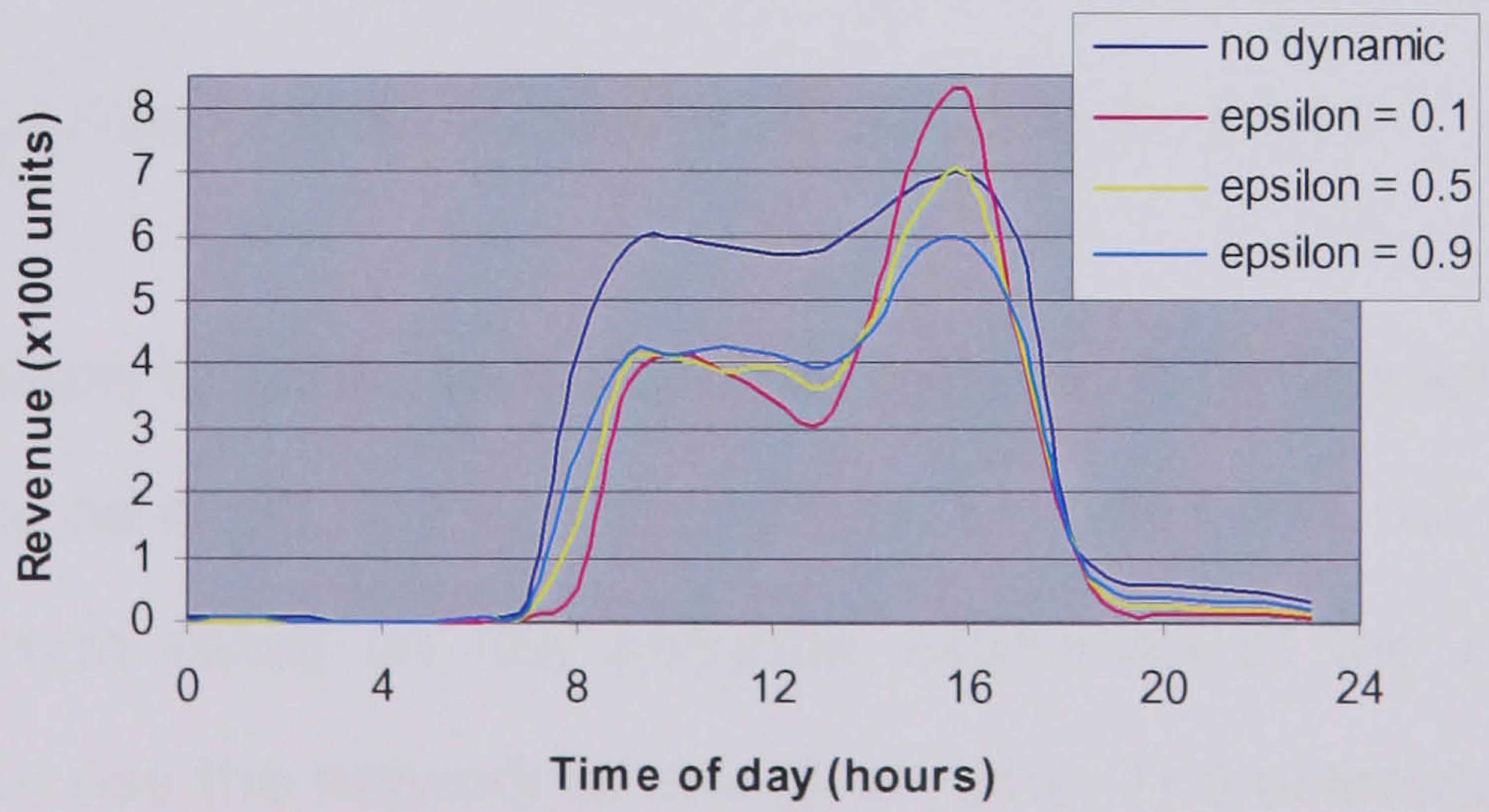

Figure 6-7 *Revenue as a function of the time of the day for* $\beta = 1.0$

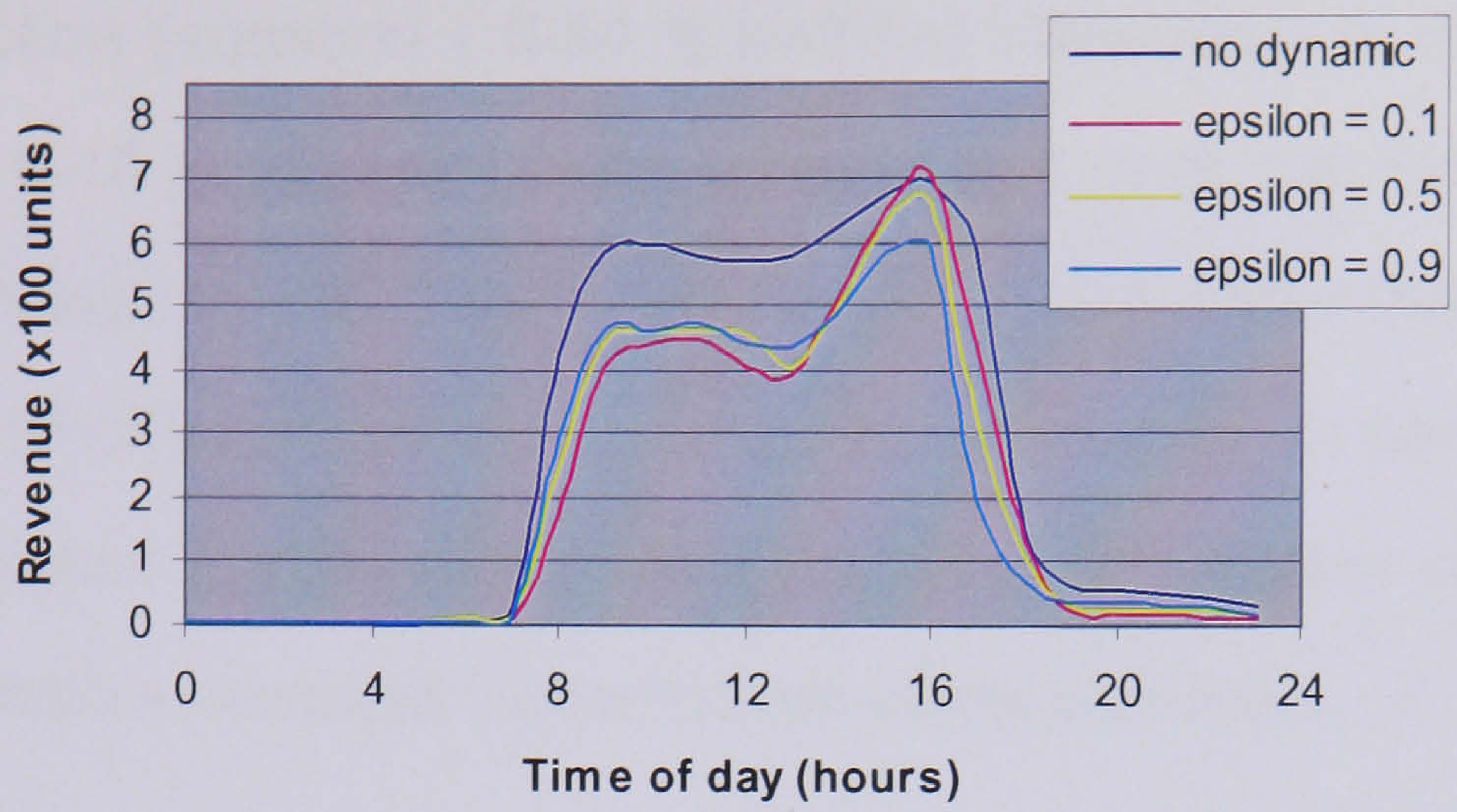

Figure 6-8 *Revenue as a function of the time of the day for* $\beta = 2.0$

During off-peak hours, however, the best match between target and actual revenue is attained, as expected, for $\varepsilon = 0.9$ for all values of demand elasticity β (Figure 6-9 and Figure 6-10).

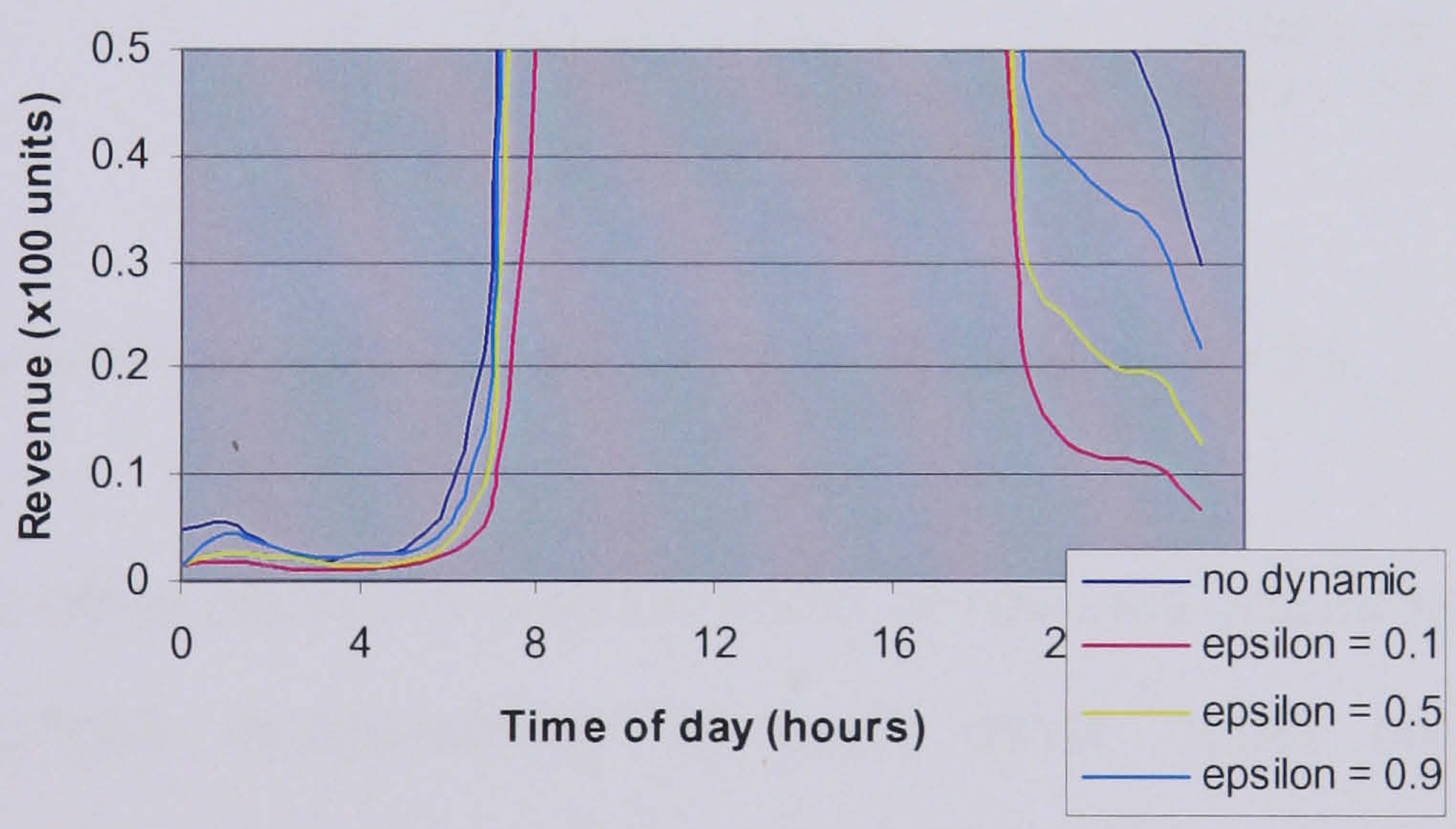

Figure 6-9 *Revenue attained during off-peak hours for* $\beta = 1.0$

The failure to generate the desired revenue is remarkably consistent and can probably be explained by three factors. On one hand, the calibration of the pricing function relies on the accurate estimation of the number of users attempting to use the network at any given time. The estimated expected load in the network (q_{fixed}) affects the pricing at two stages: the calibration of the pricing function (equation (6-14)) and the calculation of the optimal prices (equation (6-15)). Therefore, the accumulated error can become significant. This hypothesis is confirmed by the results, which show that the error in the calculation of off-peak prices is smaller than the error in peak-hour prices. A potential solution to this problem that could be investigated in the future is the use of alternative measure for estimation of network load.

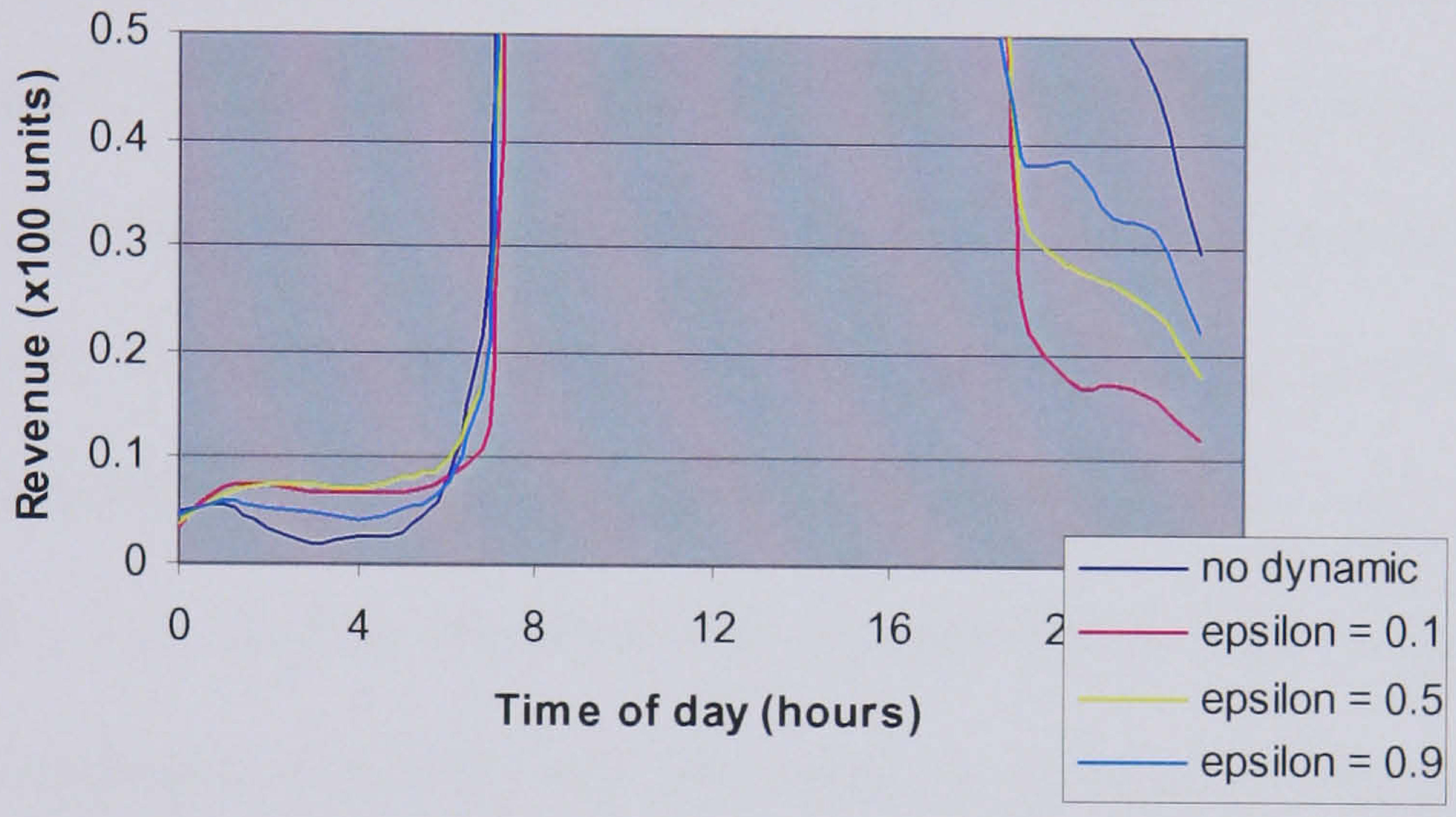

Figure 6-10 *Revenue attained during off-peak hours for* $\beta = 2.0$

On the other hand the determination of revenue attainment and capacity utilisation prices independently of each other could contribute to the cumulative error with this pricing policy. The "best" price for a particular situation, for example, the price generating a target revenue, is derived without taking into account its effect on the capacity utilisation of the network. The effect of this independent choice would be most apparent when there is a significant discrepancy in the magnitude of the revenue attainment and capacity utilisation prices, as averaging the prices using equation (6-13) will lead to a large bias in the final price.

Finally, the simplifying assumption of price linearity could [illegible]cing the affecting the effectiveness of the dynamic pricing algorithm. To test th[illegible] hypothesis the price linearity requirement would be dropped and an alternative approach to dynamic pricing discussed in the following section.

Overall, the linear revenue attainment algorithm has not performed satisfactorily in ensuring that the target revenue of the network operator is generated.

The second controlled variable in the system is the percentage of blocked calls[48], plotted as a function of the network operator's preference for revenue attainment ε in Figure 6-11. It can be seen that the percentage of blocked calls, in fact, increases as the network operator's preference for revenue attainment ε increases for both inelastic ($\beta = 1.0$) and unit elastic demand ($\beta = 2.0$). In this instance the system behaves as expected, because as network operator's preference for capacity utilisation decreases the control over network load would also decrease.

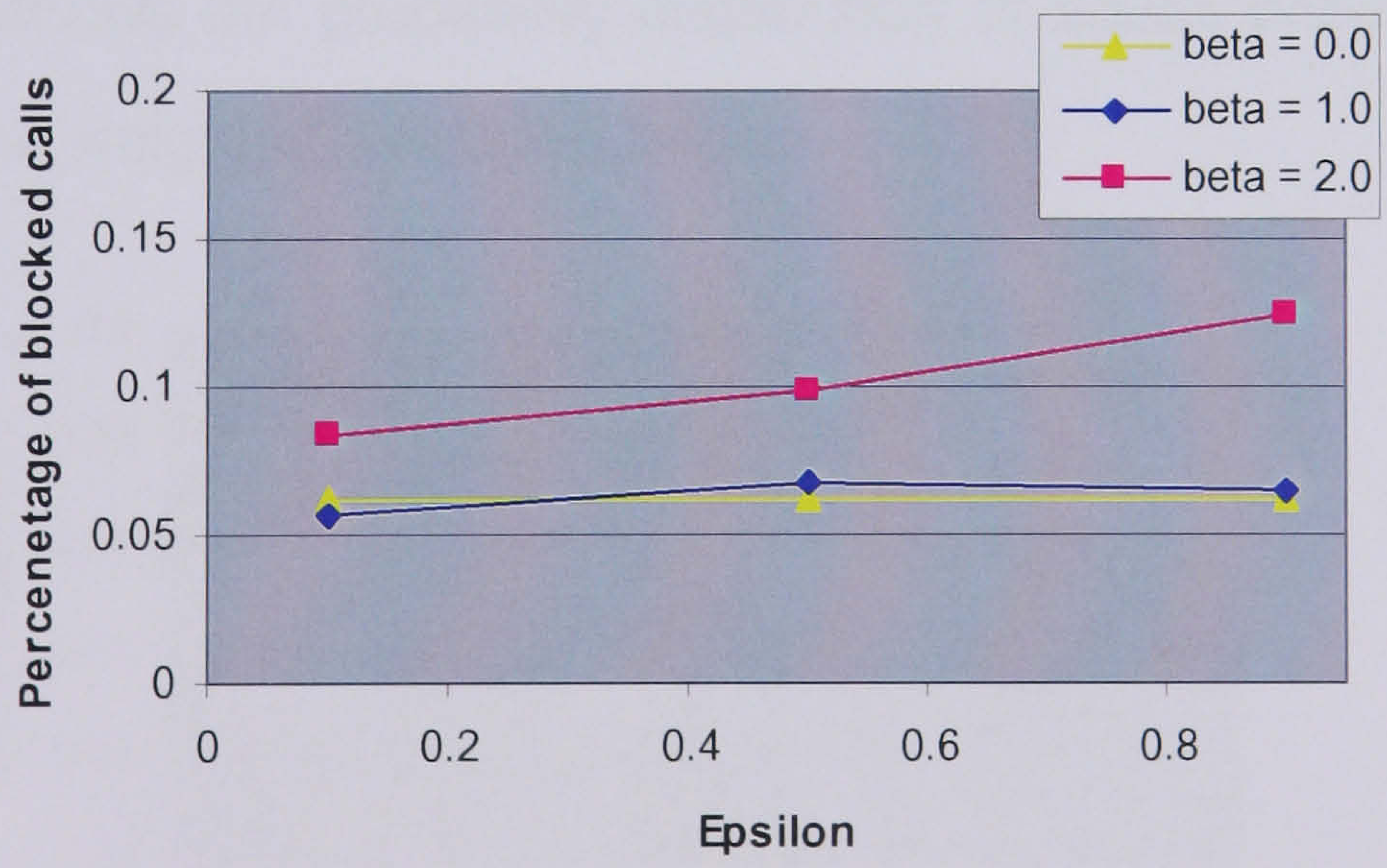

Figure 6-11 *Percentage of blocked calls with revenue attainment prices*

[48] Percentage of blocked calls is defined as the ratio of the number blocked calls over the number of successful calls in the network.

The revenue attainment pricing strategy is successful in reducing the percentage of blocked calls for inelastic demand ($\beta = 1.0$, $\varepsilon = 0.1$) by 7.5%. However, the price is not very good at controlling the blocking when demand is unit elastic ($\beta = 2.0$, $\varepsilon = 0.1$) allowing a 36% increase in the percentage of blocked calls.

While the revenue attainment strategy is not successful in controlling the revenue generated in the network and would not be considered a good choice by the network operator, it is efficient at controlling and reducing call blocking (when used with revenue attainment preference $\varepsilon = 0.1$). This strategy is also successful in generating a larger proportion of calls in the network (8% increase for inelastic demand ($\beta = 1.0$) and 35% increase for unit elastic demand ($\beta = 2.0$)). It also leads to lower average prices (up to 34% reduction) in the network compared to the scenario without dynamic pricing. Graphs and detailed discussion of these results are given in section F.4 of Appendix F.

A particular drawback with this pricing policy is highlighted by Figure 6-12, which shows the probability distribution of actual prices in the network rather than the weighted averaged price.

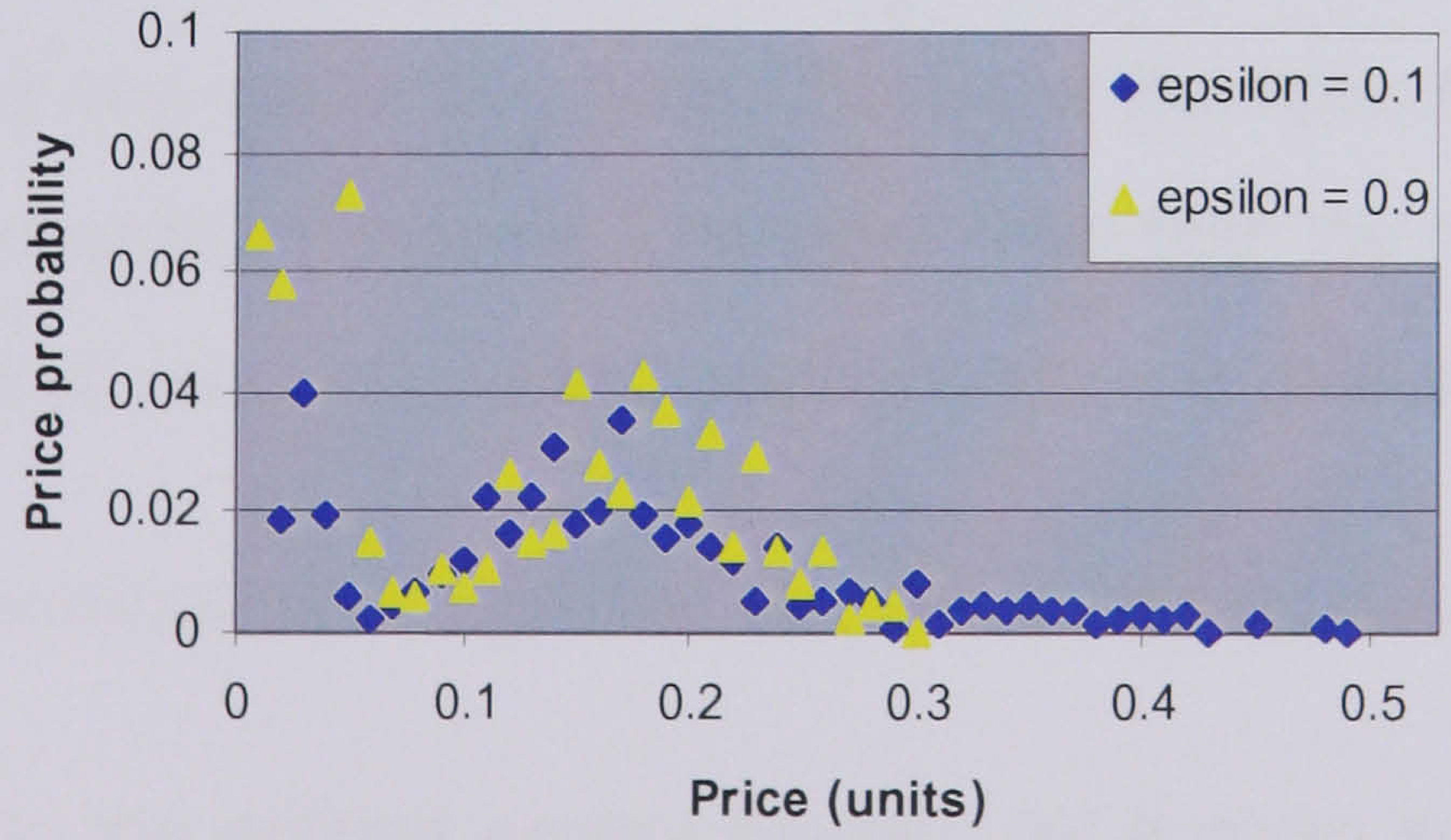

Figure 6-12 *Actual prices with revenue attainment model*

As a result of the specific way in which prices in the network are calculated it is very difficult for the network provider to control the number of prices that users see. For example, with demand elasticity $\beta = 1.0$ and revenue attainment preference $\varepsilon = 0.1$, users in the network could see up to 48 different prices, some of which are very unlikely, but nevertheless occur when the network is very busy, for instance. The total number of possible dynamic prices decreases to 30 for $\varepsilon = 0.9$ but the proliferation of prices would remain a significant problem for users. The large number of intermediate dynamic prices could cause problems for the network operator with internal accounting and billing.

6.3 Optimal Revenue Attainment Dynamic Pricing Strategy.

The dynamic pricing strategy suggested in section 6.2 assumed linearity for the dynamic price that a network operator should charge. This assumption imposes some limitations on the performance of the dynamic pricing function and ideally a network operator would be able to choose a "best" or “optimal” shape for the dynamic pricing function. In addition, the suggested revenue attainment algorithm leads to a large number of intermediate prices and would make billing and accounting very difficult (for example, see section 3.3.2.1 for a discussion of billing issues). Therefore, from both the network operator's and user's point of view, a simplified pricing function is desirable.

6.3.1 Determination of Optimal Dynamic Pricing Function.

Finding the optimal control function for a given system is equivalent to the problem of finding the optimal pricing function for the system shown in section 6.1, Figure 6-4 to steer the system into the desired state. Mathematical methodology exists in the calculus of variations, which uses Euler's equation to find optimal functions that will satisfy certain conditions (Butkov [98]). Thus calculus of variations would be a very good candidate to apply to the control problem to enable us to find the shape of the optimal pricing function. However, Pinch [108] has indicated that there is a significant difference between the description of control systems, where systems are described in terms of differential equations and the calculus of variations, which uses functionals. Therefore, in order to use the techniques from the calculus of variations to find the optimal control for a system, the statement of the problem has to be modified.

In calculus of variation terms the problem can be defined as:

$$\text{Minimise } \int_{t_0}^{t_1} \left(C - \frac{D(P(q))}{T} + \frac{q}{\tau} \right)^2 dt \qquad (6\text{-}16)$$

$$\text{subject to } \int_{t_0}^{t_1} \frac{D(P(q))}{T} P(q) \tau dt = R_{desired} \qquad (6\text{-}17)$$

t_0 - Start of optimisation time

t_1 - End of optimisation time;

C - Total network capacity;

$D(P(q))$ - Compound demand for services;

$P(q)$ - Price for calls;

q - Instantaneous network load (as number of calls);

τ - Call holding time.

This is an isoperimetric problem with fixed end points. Although in this case we chose to minimise the difference between compound system demand and available network capacity, subject to given revenue, the problem can be turned around. In that case, the aim would be to minimise the difference between the desired and current revenue, subject to a fixed capacity constraint. The dynamics of the cellular system at any intermediate optimisation time t can be expressed as:

$$\frac{dq}{dt} = \frac{D(P(q))}{T} - \frac{q}{\tau} \tag{6-18}$$

Substituting $D(P(q))$ into (6-16) and (6-17) gives the problem in terms of the change in the load $\frac{dq}{dt}$ and therefore in a calculus of variations form:

$$\text{Minimise } \int_{t_0}^{t_1} \left(C - \frac{dq}{dt} \right)^2 dt \tag{6-19}$$

$$\text{subject to } \int_{t_0}^{t_1} \left(\frac{dq}{dt} + \frac{q}{\tau} \right) P(q) \tau dt = R_{desired} \tag{6-20}$$

By letting $t = x_1$, $q = x_2$ and $\frac{dq}{dt} = u$ and introducing a further state variable x_3 with $x_3(t_0) = 0$ and $x_3(t_1) = R_{desired}$ with:

$$x_3 = \left(u + \frac{x_2}{\tau} \right) P(x_2) \tau \tag{6-21}$$

the problem is converted into an optimal control problem with fixed end points (Pinch [108] :pp 173).

It should be noted that in this definition of the problem, the load in the network is the control variable, rather than the price of the calls. This is due to the fact that we do not want to impose any restrictions on the shape of the price function, in the form of an additional differential equation. However, the price is linked to the load in the network and as the optimal load is known in advance, this will enable us to derive a function for the optimal price.

By Pontyagins maximum principle [108], in order to find the optimal control u we need to examine the function:

$$H = -(C-u)^2 + \psi_1 + \psi_2 u + \psi_3\left(u + \frac{x_2}{\tau}\right)P(x_2)\tau \qquad (6\text{-}22)$$

With the following constraints:

$$\dot{\psi}_1 = \frac{d\psi_1}{dt} = -\frac{\partial H}{\partial x_1} = 0 \qquad (6\text{-}23)$$

$$\dot{\psi}_2 = \frac{d\psi_2}{dt} = -\frac{\partial H}{\partial x_2} = -\psi_3\left(uP(x_2) + \left(u + \frac{x_2}{\tau}\right)P'(x_2)\right) \qquad (6\text{-}24)$$

$$\dot{\psi}_3 = \frac{d\psi_3}{dt} = -\frac{\partial H}{\partial x_3} = 0 \qquad (6\text{-}25)$$

The control function minimising (6-19) will be the control function maximising H and therefore we require:

$$\frac{\partial H}{\partial u} = 2C - 2u + \psi_2 + \psi_3 P(x_2)\tau = 0 \qquad (6\text{-}26)$$

This is a maximum as:

$$\frac{\partial^2 H}{\partial u^2} = -2 \le 0 \qquad (6\text{-}27)$$

The function to be optimised (H) is independent of ψ_1 and, so to find the optimal control only equations (6-24) - (6-26) need to be solved simultaneously.

The solution to equation (6-25) is straightforward and gives:

$$\psi_3 = \zeta \qquad (6\text{-}28)$$

ζ - a constant.

To find ψ_2 we use (6-26) to get:

$$\frac{d}{dt}\left(\frac{\partial H}{\partial u}\right) = -2u' + \dot{\psi}_2 = 0 \tag{6-29}$$

Substituting into (6-24) we get:

$$2u' = \zeta\left(uP(x_2) + \left(u + \frac{x_2}{\tau}\right)P'(x_2)\right) \tag{6-30}$$

Solving (6-30) with respect to u gives[49]:

$$u = e^{\frac{1}{2}t\zeta(P+P')}C_1 - \frac{P'x_2}{(P'+P)\tau} \tag{6-31}$$

C_1 - Integration constant.

To find the value of C_1 we assume that at time $t = 0$ the system has a load of q_0. In this case:

$$C_1 = q_0 + \frac{P'x_2}{(P'+P)\tau} \tag{6-32}$$

Therefore, the optimal control for this particular system as a function of time is:

$$u = \frac{-x_2P' + e^{\frac{1}{2}t\zeta(P+P')}\left((P+P')\tau q_0 + P'x_2\right)}{(P'+P)\tau} \tag{6-33}$$

Using equation (6-20) will enable us to find ζ and thus the optimal load in the network as a function of time. The optimisation of the revenue occurs at independent small time intervals within each hour and, so for each optimisation subinterval $t_0 = 0$ and $t_1 = t_1$.

The constraint equation to be solved becomes:

$$\int_0^{t_1}\left(u + \frac{x_2}{\tau}\right)P\tau dt = R_{desired} \tag{6-34}$$

[49] For simplification the dependence of the price $P(x_2)$ on the load x_2 will not be written explicitly.

$$\frac{P}{(P+P')^2\zeta}\left(\begin{array}{l}2\left(-1+e^{\frac{1}{2}(P+P')t_1\zeta}\right)(P+P')\tau q_0+\\ x_2\left(2\left(-1+e^{\frac{1}{2}(P+P')t_1\zeta}\right)P'+P(P+P')t_1\zeta\right)\end{array}\right)=R_{desired} \qquad (6\text{-}35)$$

The value of ζ cannot be derived analytically and therefore, the equation has to be solved numerically using the technique of exhaustive enumeration [107] by testing all possible numerical combinations of P and P'. To ensure that the solutions of equation (6-35) are feasible, it has to be solved simultaneously with equation (6-33), as the load in the network has to be kept as close to the optimal load as possible.

Returning to the variables in the original problem gives:

$$\frac{qP+e^{\frac{1}{2}t\zeta(P+P')}((P+P')\tau q_0+P'q)}{(P'+P)\tau}=\frac{D(P(q))}{T} \qquad (6\text{-}36)$$

$$\frac{P}{(P+P')^2\zeta}\left(\begin{array}{l}2\left(-1+e^{\frac{1}{2}(P+P')t_1\zeta}\right)(P+P')\tau q_0+\\ q\left(2\left(-1+e^{\frac{1}{2}(P+P')t_1\zeta}\right)P'+P(P+P')t_1\zeta\right)\end{array}\right)=R_{desired} \qquad (6\text{-}37)$$

Equations (6-36) and (6-37) provide a very general solution for finding optimal prices for a control system defined by equations (6-16) and (6-17). A specific solution to the system for any particular set of initial conditions can in turn be found by using information relating to the distinct system the network operator is using.

6.3.2 Network Specific Optimal Pricing Solution.

We have a general optimal pricing model, that when applied to a specific network, will allow the network operator to find the optimal dynamic prices for any type of demand as a function of price and time of day fluctuations in the network. To find a pricing solution specific to the network used for simulations, the corresponding variables are given their numerical values. It was assumed that the average call holding time is $\tau = 1.2$ minutes, price updates occur every 5 minutes and the price remains constant within the subintervals. As a result of this assumption, the price in the subinterval will be determined by the load in the network at time $t = 0$, which by definition is q_0 and $t_1 = 1$ and $T = 1$.

By taking into account these assumptions a solution to the general revenue attainment problem defined in (6-16) and (6-17) will require the numerical solution of:

$$\left| \begin{array}{l} D(\beta, P) = \dfrac{\left(P + 1.2e^{\frac{1}{2}(P+P')\zeta}(1.83P' + P) \right) 0.834 q_0}{P' + P} \\ R_{desired} = \dfrac{P q_0 \left(P(-2.4 + P\zeta) + (-4.4 + P\zeta)P' + 2.4e^{\frac{1}{2}(P+P')\zeta}(P + 1.83P') \right)}{(P + P')^2 \zeta} \end{array} \right. \tag{6-38}$$

A Matlab program for finding the optimal prices for the exponential demand function discussed in section 4.1.2 is shown in Appendix D.

An interesting property of the specific optimal solution to the revenue attainment problem is that it involves both the price of the calls P and the derivative of the price P'. However, the demand function developed in section 4.1.2 does not involve the absolute change in the price P'. This means that in this case, the solutions we will find will not be optimal. To ensure full optimality the model for demand will have to be enhanced to take into account the expected reaction of users to the change in price. In this thesis it will be assumed that the derivative of the price $P' = 0$ and the simplified model for finding the optimal prices becomes:

$$\begin{aligned} D(\beta,P) &= \left(0.834 + e^{\frac{P\zeta}{2}}\right) q_0 \\ R_{desired} &= \frac{2.4 q_0 \left(-1 + e^{\frac{P\zeta}{2}} + 0.47 P\zeta\right)}{\zeta} \end{aligned} \qquad (6\text{-}39)$$

Using (6-39) with the exponential network demand discussed in section 4.1.2 and the desired revenue without dynamic pricing plotted in Figure 6-5, the optimal dynamic prices given in Table 6-1 and plotted in Figure 6-13 are derived. It can be seen that as the elasticity of demand β increases, the optimal prices in the network decrease on average by 10%. Overall as the load in the network increases the optimal dynamic prices also increase.

To reduce the number of potential prices and simplify the pricing tariffs it is suggested that the prices would only change for load increases of 10%.

Network utilisation (% busy channels)	**Optimal price**	
	$\beta = 1.0$	$\beta = 2.0$
1.75	0.02	0.01
2.1	0.02	0.01
2.45	0.02	0.01
2.8	0.02	0.01
5.25	0.02	0.01
7.7	0.04	0.03
10.85	0.01	0.01
23.8	0.03	0.02
30.1	0.06	0.06
40.95	0.03	0.03
45.5	0.03	0.03
52.15	0.03	0.03
53.2	0.23	0.22
56	0.02	0.02

75.6	0.15	0.27
78.4	0.15	0.27
80.85	0.31	0.28
82.95	0.17	0.17
83.65	0.33	0.29
83.65	0.33	0.29
85.4	0.04	0.04
91.7	0.43	0.34
109.9	0.5	0.38
113.75	0.55	0.4

Table 6-1 *Optimal dynamic prices plotted against network utilisation.*

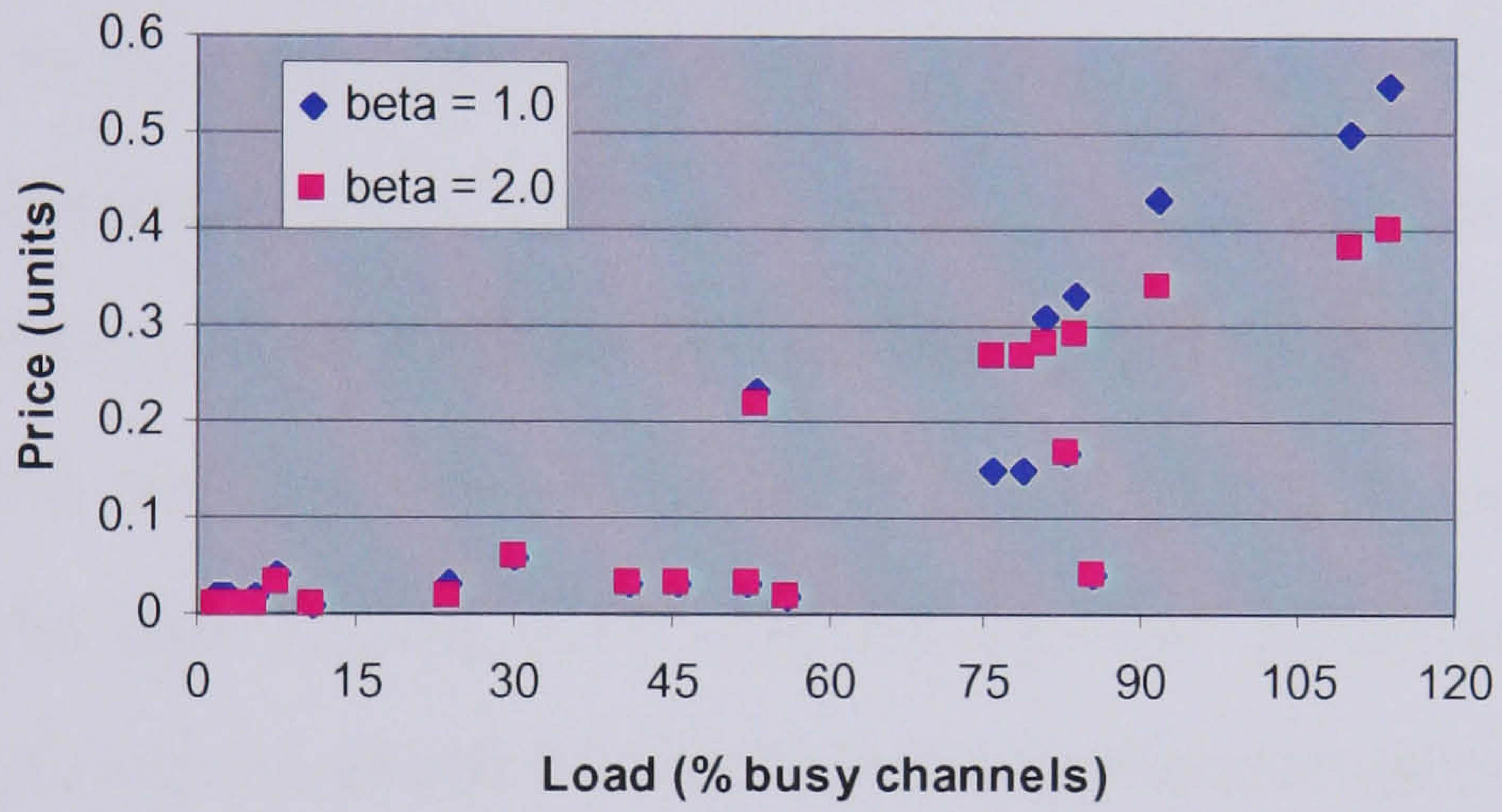

Figure 6-13 *Optimal dynamic revenue attainment prices*

The prices given in Table 6-1 were derived on the basis of the system behaviour in real time, and this has two consequences. On one hand, the optimal prices for some states of the network cannot be found such as at loads between [57%,75%] as, due to demand variation, real time data is not available. In this case, the optimal price will have to be estimated using, for example, regression. On the other, for some states of the network more than one optimal price was found, for example, the load between 0% and 10%. These optimal prices can potentially be different, in which case the average of the two prices will be taken as the optimal price.

Averaging the prices, however, can lead to a significant problem. The optimal prices divide naturally into two separate groups for the same load in the network for all values of demand elasticity β, depending on the time at which this load occurred. This division is due to the different 'inherent' peak and off-peak prices in the network. The optimal prices for loads that occur during peak hours are significantly higher than the optimal prices for the same load during off-peak hours. This can be seen in Table 6-1 for loads in the interval $[55\%,85\%]$ with the optimal price fluctuation within the interval of $[0.02,0.33]$ units (demand elasticity $\beta = 1.0$).

Averaging the optimal price in this case will lead to an overall reduction/increase in the optimal prices that will decrease the optimality of the pricing strategy. This problem can be overcome by allowing peak and off-peak tariffs in the network depending on the time of day a load occurs. Plotted in Figure 6-14 and Figure 6-15 are the optimal dynamic prices for demand elasticity $\beta \in [1.0,2.0]$, which will be tested with the simulation[50].

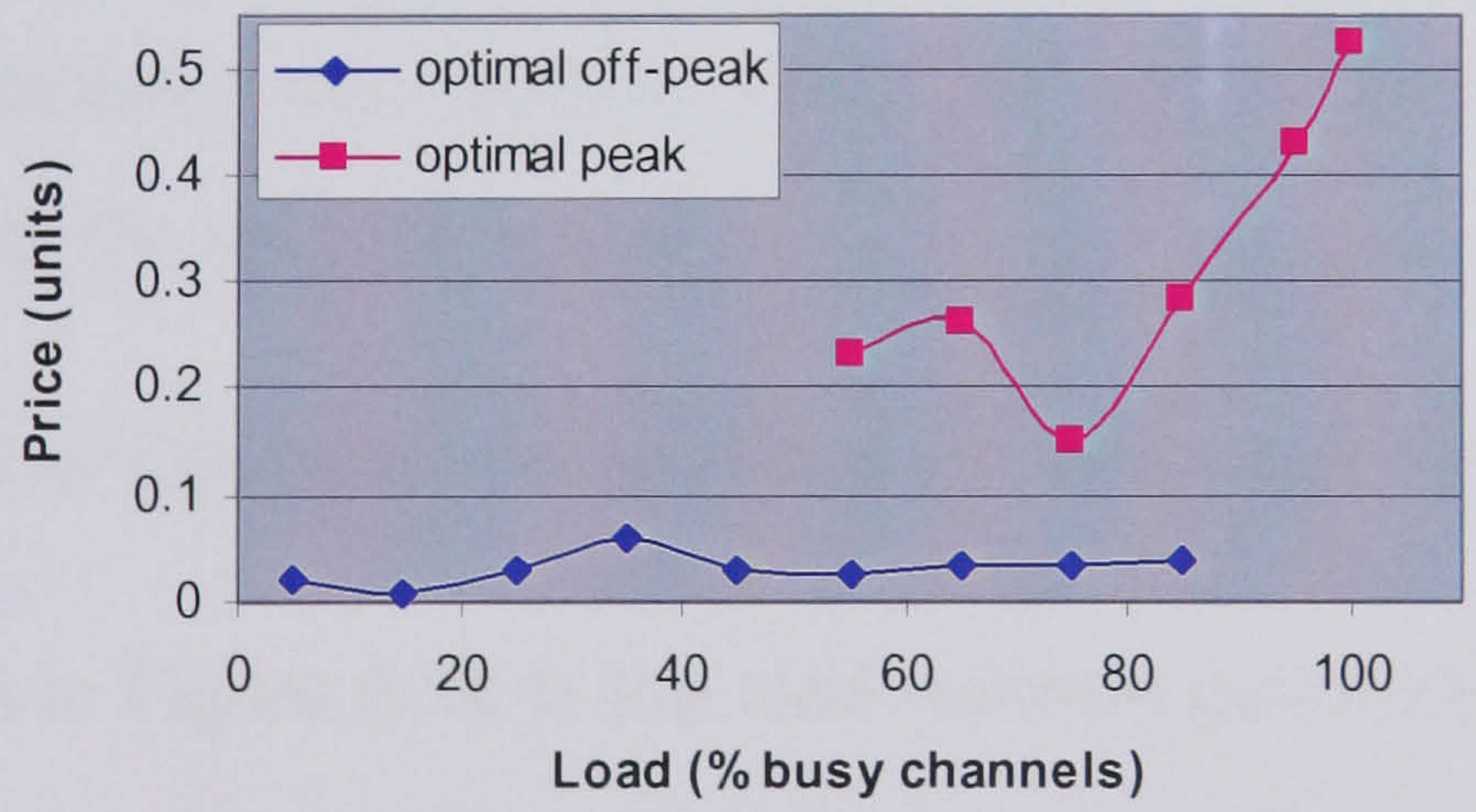

Figure 6-14 *Optimal dynamic revenue attainment prices for* $\beta = 1.0$

[50] Missing prices were estimated using Microsoft Excel 97 SR -1 built in polynomial and power functions (peak and off peak hours respectively).

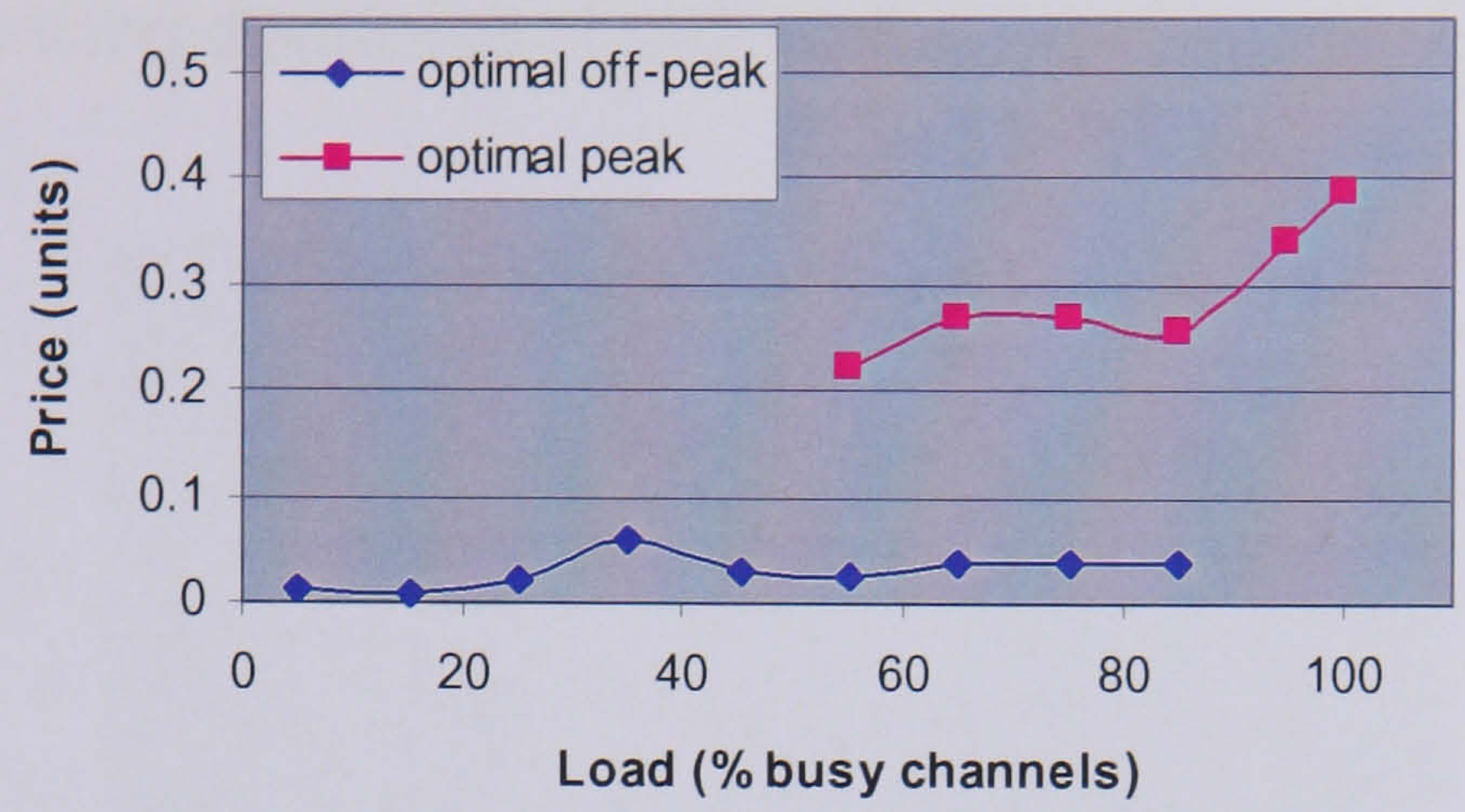

Figure 6-15 *Optimal dynamic revenue attainment prices for* $\beta = 2.0$

6.4 Simulation Results for Optimal Revenue Attainment Pricing.

The efficiency of the optimal dynamic pricing as a control to ensure the generation of a target revenue and reduction of call blocking was tested using the OPNET™ simulation developed in section 5.2. The results from the simulations were compared with the corresponding results from the linear revenue attainment model. The network operator preference for revenue and capacity attainment was set at $\varepsilon = \phi = 0.5$. This choice gives equal weight to revenue attainment and capacity utilisation, and is perhaps the most natural alternative for the network operator.

6.4.1 Effect of Optimal Pricing Policy on Revenue Generation.

Plotted in Figure 6-16 is the total revenue generated in the network with the optimal dynamic pricing strategy, compared to the revenue attainment strategy. It can be seen that the generated revenue with the optimal dynamic pricing is very close to the total desired revenue (in this case this is the revenue generated without dynamic pricing *i.e.* demand elasticity $\beta = 0.0$). This is true for both inelastic ($\beta = 1.0$) and unit elastic ($\beta = 2.0$) demand. The optimal dynamic pricing performs significantly better than the linear revenue attainment pricing discussed in section 6.2.

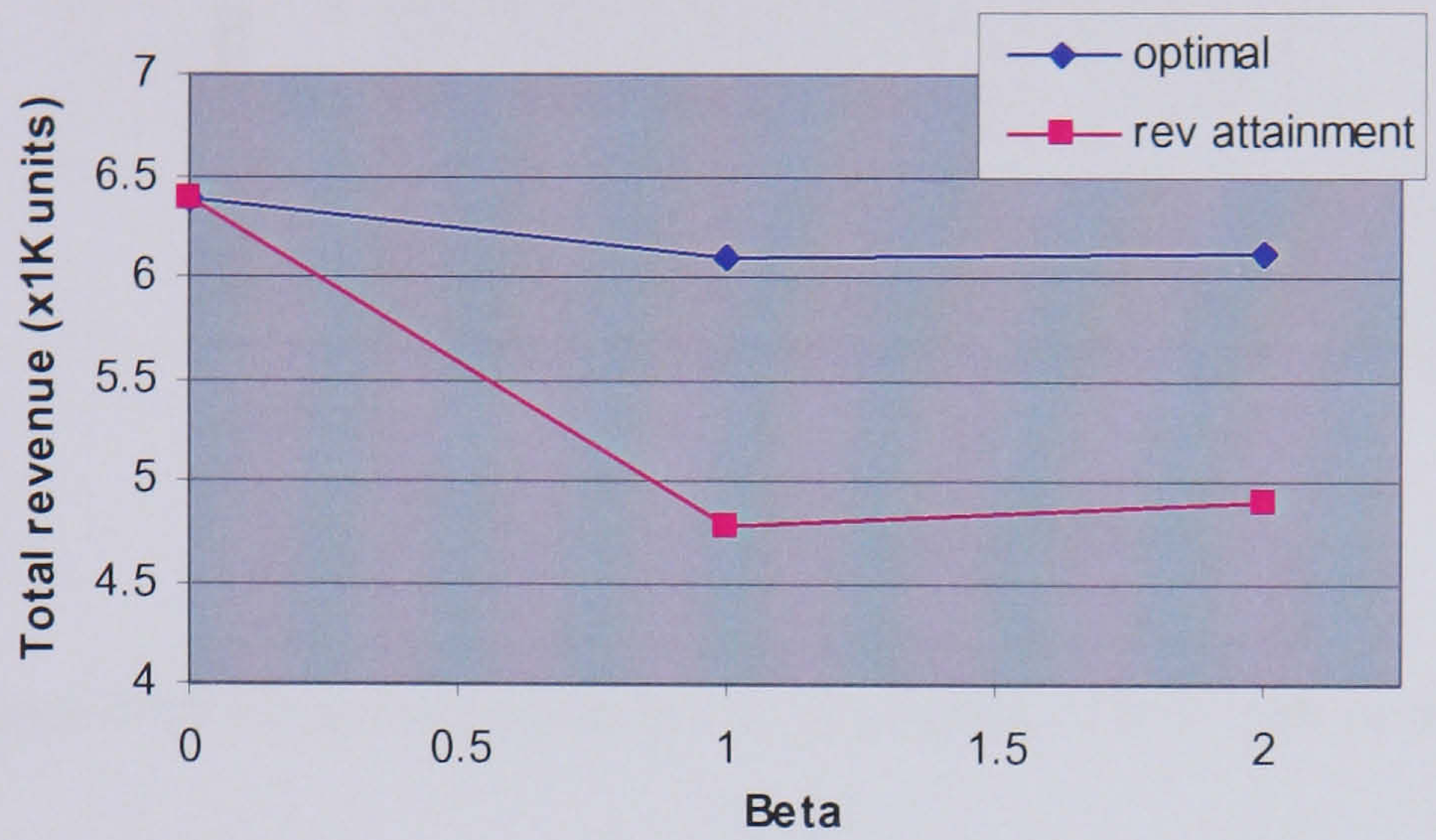

Figure 6-16 *Total revenue generated with optimal and linear revenue attainment pricing*

Although it performs better that the linear revenue attainment pricing, the revenue generated with the optimal pricing is 4% below the target level (set as the revenue without dynamic pricing for $\beta = 0.0$), regardless of the demand elasticity. Although this is not a significant reduction, in the long run it could have more significant effects. The discrepancy can be explained in part by the averaging of the prices to reduce the overall number of prices in the network. In addition, the model used for the derivation of the optimal prices is analogue by nature, while the optimal prices are discrete. Further research would be necessary to establish the degree of error introduced by the rounding of prices in the process of making them discrete.

Comparing the revenue generated at different times of the day shows that the target revenue was generally attained when the load in the network was relatively small and demand was inelastic ($\beta = 1.0$) (see Figure 6-17 and Figure 6-18).

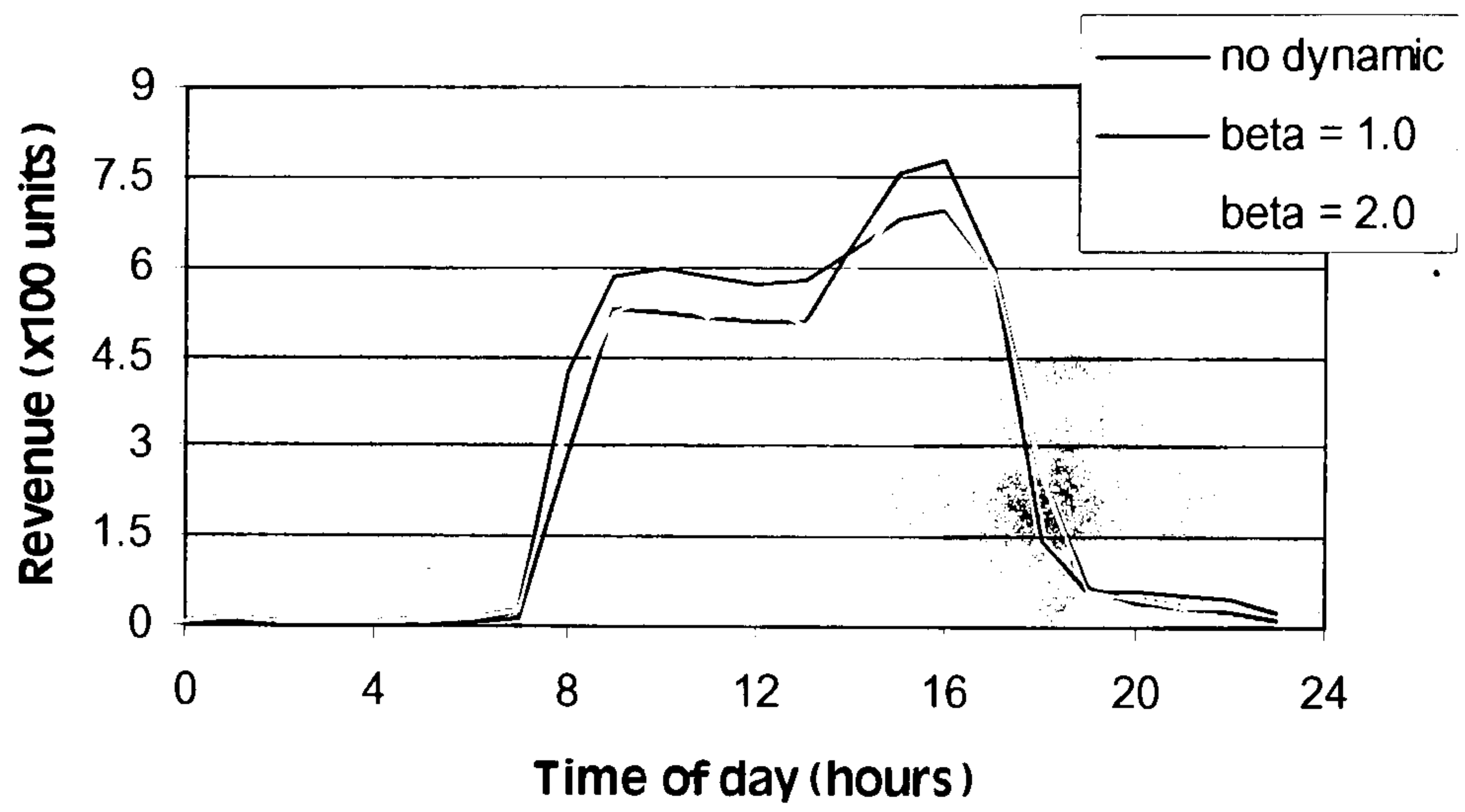

Figure 6-17 *Revenue generated as a function of time with optimal dynamic prices*

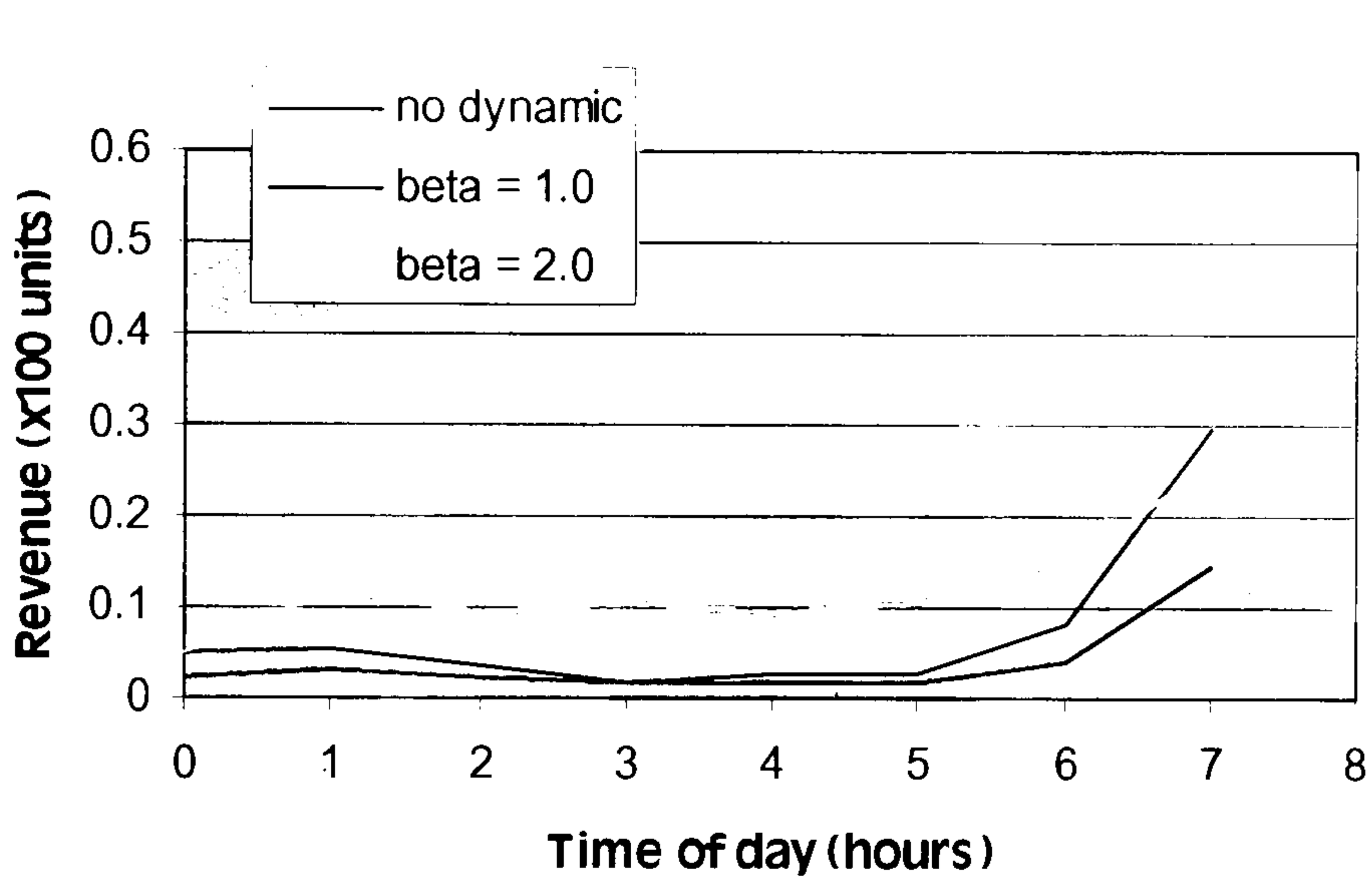

Figure 6-18 *Revenue generated as a function of time with optimal dynamic prices (small network loads)*

However, the target hourly revenue was not attained when either the network was very busy or demand elasticity β was increased. This problem was also encountered with the linear revenue attainment pricing policy. Nevertheless, the total revenue generated is not significantly different from the total desired revenue and, therefore, the optimal pricing policy is successful in controlling the cellular system to satisfy a given revenue requirement, through the setting of prices on a dynamic basis.

6.4.2 Effect of Optimal Pricing Policy on Call Blocking.

The second controlled variable of interest is the percentage of blocked calls. The effect of the optimal pricing on the total number of blocked calls in the network can be seen in Figure 6-19. The optimal pricing function leads to a 4% reduction in the percentage of blocked calls (for inelastic demand ($\beta = 1.0$)), compared to the percentage of blocked calls without dynamic pricing. As demand becomes unit elastic ($\beta = 2.0$), the percentage of blocked calls increases by 25%, compared to the blocking without dynamic pricing. This increase is due to the increased demand elasticity and was also observed with revenue attainment strategy. The failure of the optimal pricing function to control call blocking with unit elastic demand (demand elasticity $\beta = 2.0$) could be due to other parameters in the system such as the frequency of the price update interval.

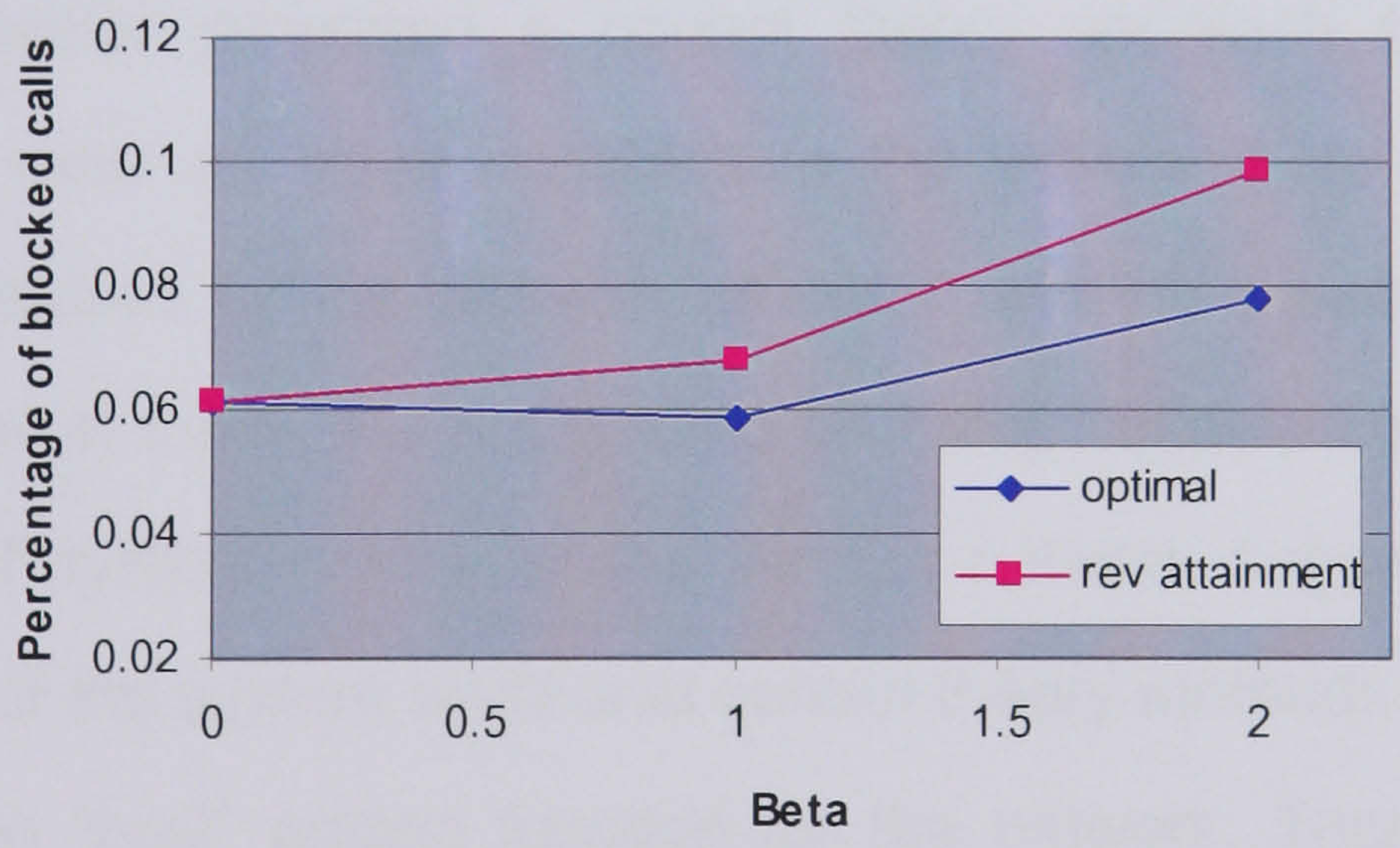

Figure 6-19 *Percentage of blocked calls with optimal and revenue attainment pricing*

However, overall the optimal pricing function performs better than the linear revenue attainment strategy, regardless of demand elasticity, and therefore, the optimal pricing function is significantly better at controlling the percentage of blocked calls.

The effect of the optimal dynamic pricing on other parameters in the network can be seen in section F.5 (Appendix F). The optimal dynamic pricing strategy stimulates demand during off-peak hours and controls demand during peak hours (Figure F-26), increases the overall number of successful calls (Figure F-28) and leads to a reduction in the average price in the network (Figure F-29).

In summary, the optimal dynamic pricing strategy is significantly better than the linear revenue attainment pricing strategy at attaining the target revenue and limiting the percentage of blocked calls, while increasing the number of successful calls and decreasing the average price in the network. Therefore, the optimal dynamic pricing strategy is the control function recommended to network operators.

6.5 Chapter Summary.

This chapter proposed a control theory approach to dynamic price setting. The most sensitive variables in the system were identified as the revenue generated by the network operator and the percentage of blocked calls. As a result these are recognised as being suitable controlled variables and a control system model for the cellular network suggested. Due to the non-linearity of the system traditional control theory methods were not suitable for finding the "best" pricing function for the network. Therefore, alternative approaches were suggested.

First a linear revenue attainment pricing model was developed taking into account the preference of the network operator for revenue attainment ε or maximum capacity utilisation ϕ. Results from the simulation showed that the total revenue generated in the network did not match the network provider's desired revenue, when defined as the revenue generated in the network without dynamic pricing (see Figure 6-6 or Figure 6-7). This was particularly evident during peak hours, which decreases the utility of revenue attainment pricing from network provider's point of view. This discrepancy was attributed partly to the error in estimation of the expected number of users in the network from which the prices are derived, and partly to the independent choice of optimal prices for revenue attainment and capacity utilisation. The linear pricing function, however, was efficient at controlling call blocking (Figure 6-11), in particular for inelastic demand ($\beta = 1.0$). A specific problem with the revenue attainment strategy is the fact that it leads to users seeing up to 40 different dynamic prices, which is difficult both from the network operator and user's points of view.

Then the assumption of linearity for the revenue attainment pricing function was dropped and an optimal shape for the dynamic pricing function derived by modification of the revenue attainment model, using calculus of variations and control theory. A comprehensive model for the determination of optimal dynamic price setting strategies for any type of demand assumptions was presented. Simulation results showed that the optimal dynamic pricing is very successful at generating the desired revenue for the network operator for different values of the demand elasticity β (Figure 6-16). In addition, it led to a decrease in the probability of call blocking and an overall increase in the total number of calls completed in the network (Figure 6-19). This shows that the optimal pricing strategy is successful at controlling the network into the desired state and this makes it a very attractive dynamic pricing policy.

Chapter 7

The previous chapters presented three different approaches to dynamic pricing and evaluated their effectiveness from the point of view of both network operators and users. This chapter investigates the expected effect of the *ad hoc*, the linear revenue attainment and optimal revenue attainment pricing strategies on network operators' market share. The aim, from a network operator's point of view, is to identify the optimal pricing strategy that will maximise the size of the user database.

7.1 Effect of Dynamic Pricing on Network Operator's Market Share

The effect of the different dynamic pricing strategies on the network operator's market share can be estimated using the model suggested in chapter 3 equation (3-2). The suggested mathematical model compares the benefits of the users with two different pricing strategies and takes into account the value for money users receive from the service, the total number of calls completed in the network and the revenue generated from the network. It is assumed that the users will chose their service provider on the basis of the benefits received from the respective pricing strategies.

User benefit derived from the dynamic pricing strategies discussed in this thesis will be compared to the user benefit with the traditional non-dynamic pricing strategy. The graphs below show the expected proportion of users in the market preferring a dynamic pricing strategy. The loyalty of the users (m)

will be taken as equivalent for both service providers. This scenario could exist when a service provider offers both dynamic and traditional non-dynamic tariffs, in which case the model aims to predict the proportion of users who will take up the respective tariffs.

7.1.1 Effect of Ad Hoc Dynamic Pricing on Network Operator's Market Share

From Figure 7-1 we can see that if network operator were to offer competition driven *ad hoc* pricing then the linear pricing tariff could expect to capture 60% of the market, irrespective of the elasticity of user demand.

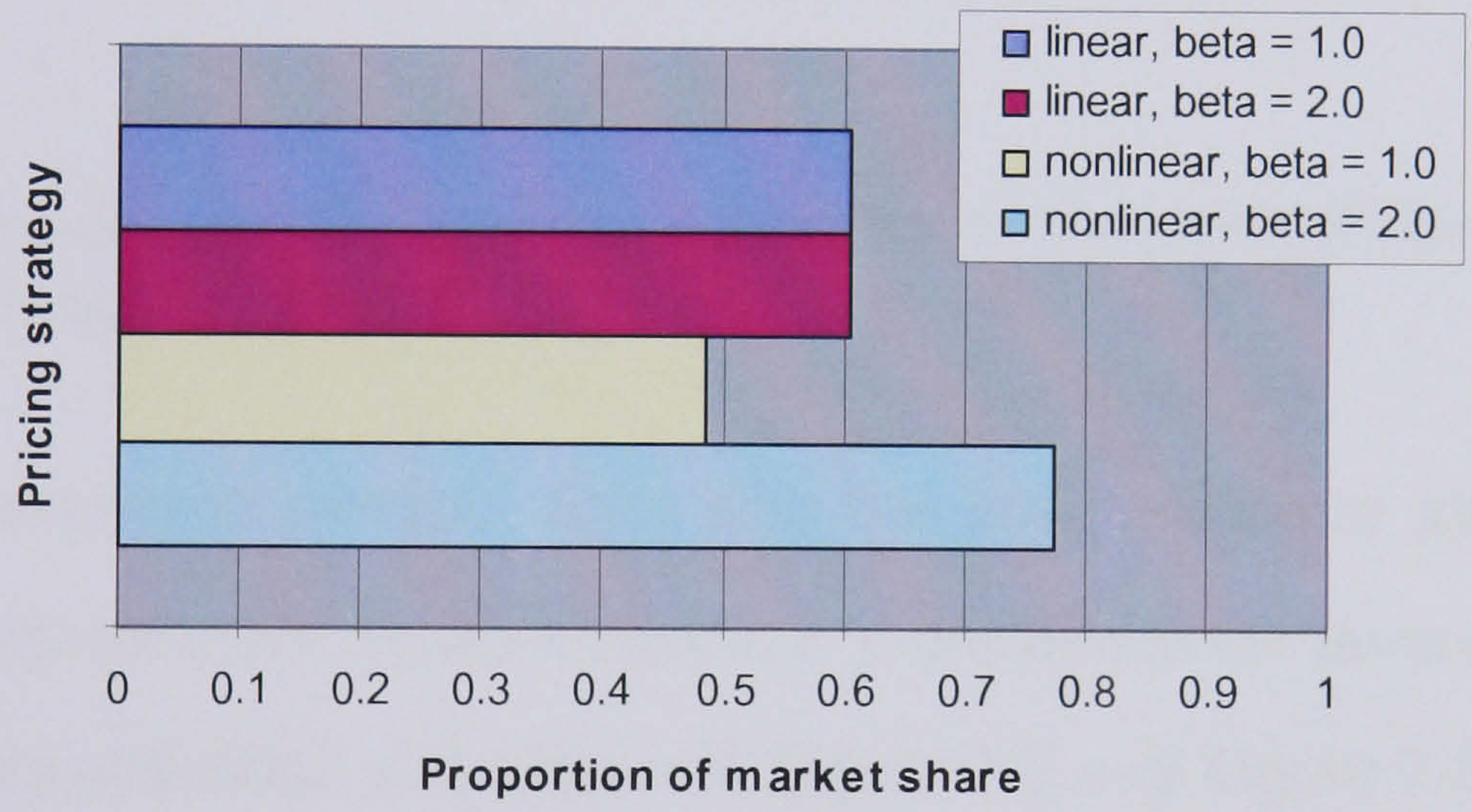

Figure 7-1 *Expected proportion of market share with competition driven ad hoc pricing*

Although the linear *ad hoc* function offers higher prices, it also leads to a significantly lower probability of call blocking which increases the "value for money component " of user benefit. In addition, the network operators can invest a significant proportion of increased revenue into advertisement and special promotions, which would attract new users[51].

[51] This prediction is based on the assumption that the proportion of people deterred by higher prices would be equal to the proportion of users attracted by the higher QoS, advertising and special promotions. This would have to be confirmed by further market research.

However, with a non-linear pricing function and inelastic demand (demand elasticity $\beta = 1.0$) the network operator could expect to capture only 48% of the market. This is due to the significantly reduced revenue and the effect this will have on the ability of the network operator to promote its services. The situation changes, though, as the elasticity of demand increases and demand becomes unit elastic (demand elasticity $\beta = 2.0$). Lower prices and significantly higher number of calls in the network outweigh the negative effects of call blocking and reduced revenue and attract a larger proportion of the market (78%). However, the retention of this market share will be a significant problem if the provided QoS is perceived as unsatisfactory by users.

7.1.2 Effect of Linear Revenue Attainment Dynamic Pricing on Network Operator's Market Share

The expected market share with the linear revenue attainment pricing strategy with different network operator preferences for revenue attainment ε and capacity utilisation ϕ is shown in Figure 7-2 and Figure 7-3.

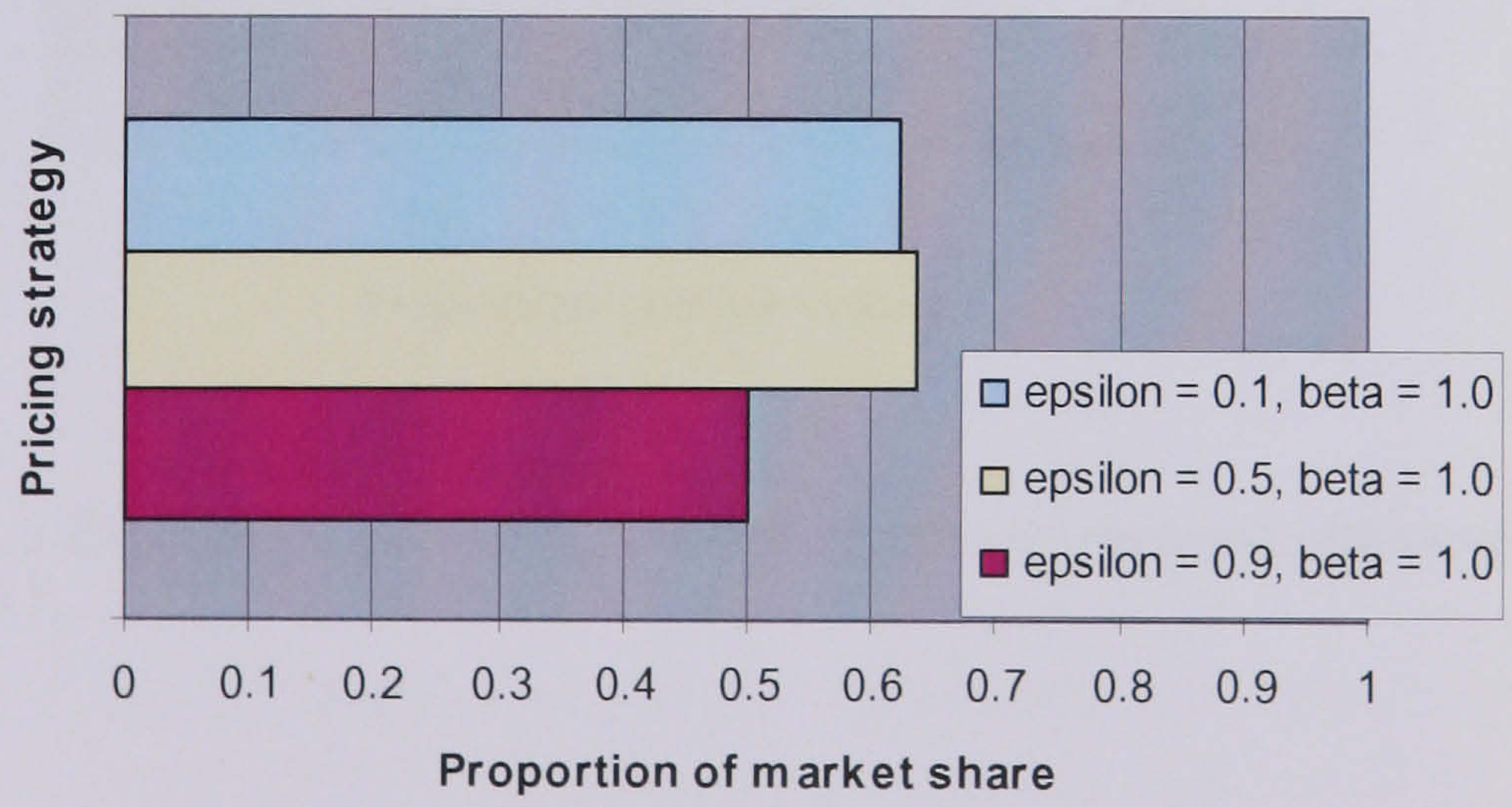

Figure 7-2 *Expected proportion of market share with revenue attainment pricing policy ($\beta = 1.0$)*

In this case, if demand is inelastic (demand elasticity $\beta = 1.0$) the best strategy for the network is equal preference for revenue attainment and capacity utilisation ($\varepsilon = \phi = 0.5$). This policy is predicted to gain 63% of the market, followed closely by the revenue attainment pricing strategy giving preference to capacity utilisation ($\varepsilon = 0.1$) with 61% of the market.

However, if demand is unit elastic (demand elasticity $\beta = 2.0$), the best strategy for capturing maximum market share is the capacity utilisation strategy with small revenue attainment preference $\varepsilon = 0.1$. This strategy would ensure 80% of the market share, compared to an equal preference for revenue and capacity attainment potentially capturing 72% of the market. This can be explained by the relative increase in the price sensitivity of users as demand elasticity β increases, which will manifest itself in a higher proportion of users giving more weight to lower prices.

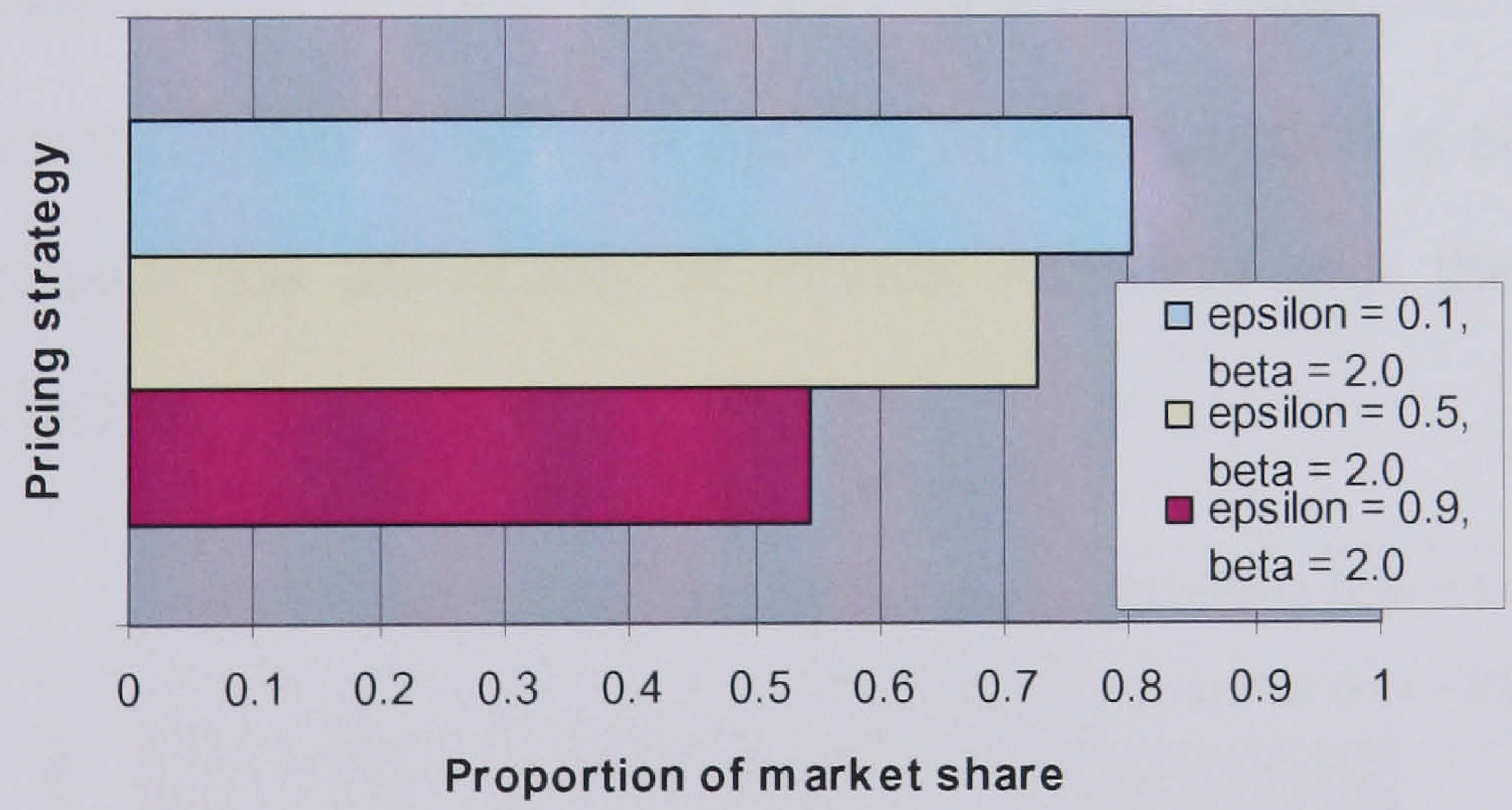

Figure 7-3 *Expected proportion of market share with revenue attainment pricing policy ($\beta = 2.0$)*

In both cases the pricing strategy giving highest preference to revenue attainment ($\varepsilon = 0.9$) performs worst capturing only 50% and 52% of the potential market when competing with a service provider offering non-dynamically priced services. This may be explained by the fact that although the average prices offered with the revenue attainment pricing strategy are lower than the average non-dynamic prices, the amount of revenue generated is also lower and does not offer adequate leverage to the network operator for service promotion.

7.1.3 Effect of Optimal Revenue Attainment Dynamic Pricing on Network Operator's Market Share

The proportion of market share that a service provider can expect to capture with the optimal dynamic pricing strategy is plotted in Figure 7-4. It is around 60% with both inelastic (demand elasticity $\beta = 1.0$) and unit elastic (demand elasticity $\beta = 2.0$) demand. This is less then the predicted market share with the non-linear *ad hoc* pricing function with unit elastic demand ($\beta = 2.0$) or the linear pricing function with low or medium preference for revenue attainment ($\varepsilon = 0.1$ or $\varepsilon = 0.5$, $\beta = 2.0$). However, the QoS users receive from the network with the optimal pricing function is significantly better and as a result the probability of market retention with the optimal pricing function is higher.

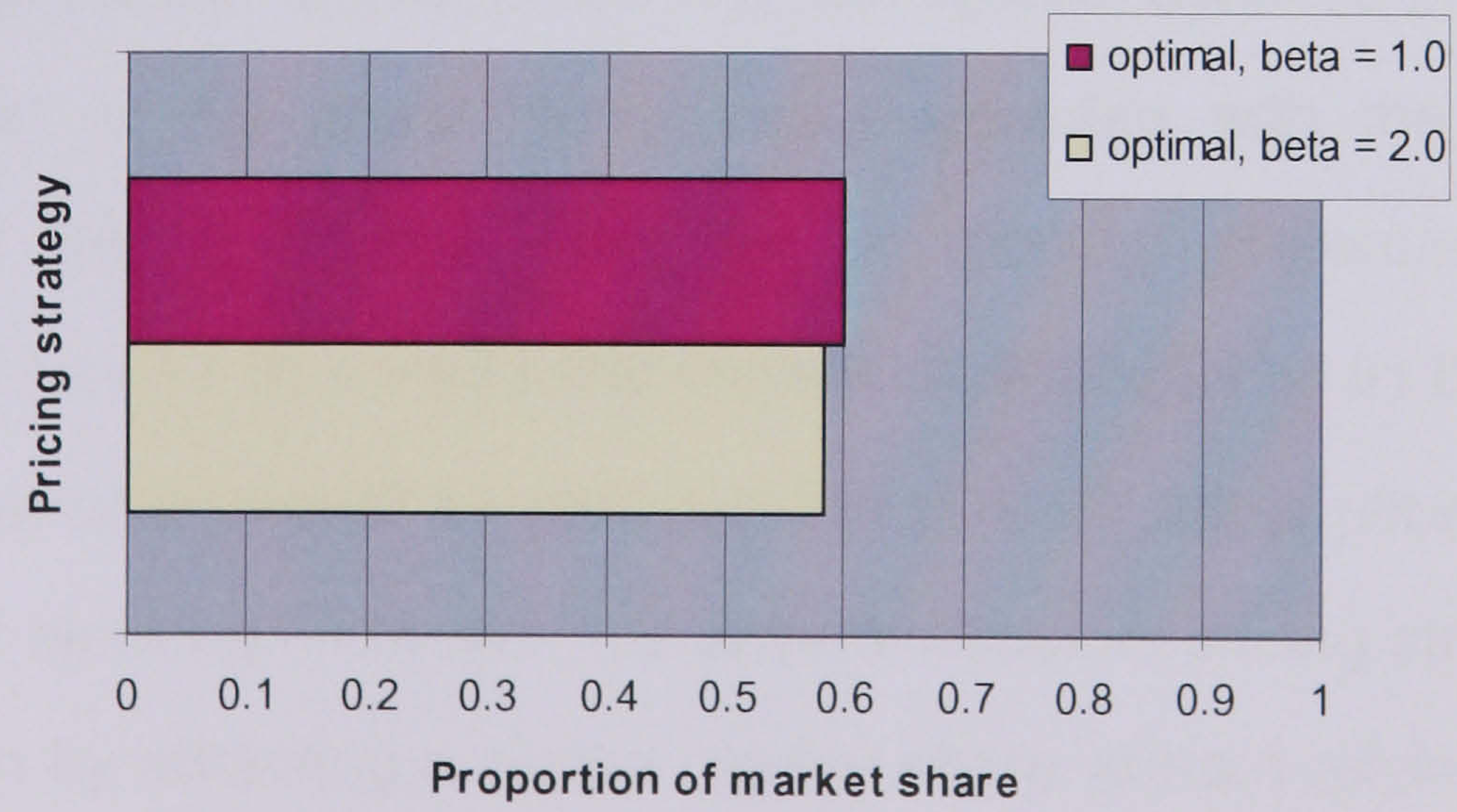

Figure 7-4 *Expected proportion of market share with optimal pricing policy*

The predicted market share with the optimal dynamic pricing function is very similar to the proportion of users expected with the linear *ad hoc* competition driven pricing strategy (demand elasticity $\beta = 1.0$ and $\beta = 2.0$) (see Figure 7-5). This could be explained by the proportion of users attracted by promotions and advertising generated by the additional revenue in the case of the *ad hoc* competition driven strategy. However, the reliance on this strategy is risky from a network operator's point of view because it depends on a random factor, in this case the success of an advertising campaign.

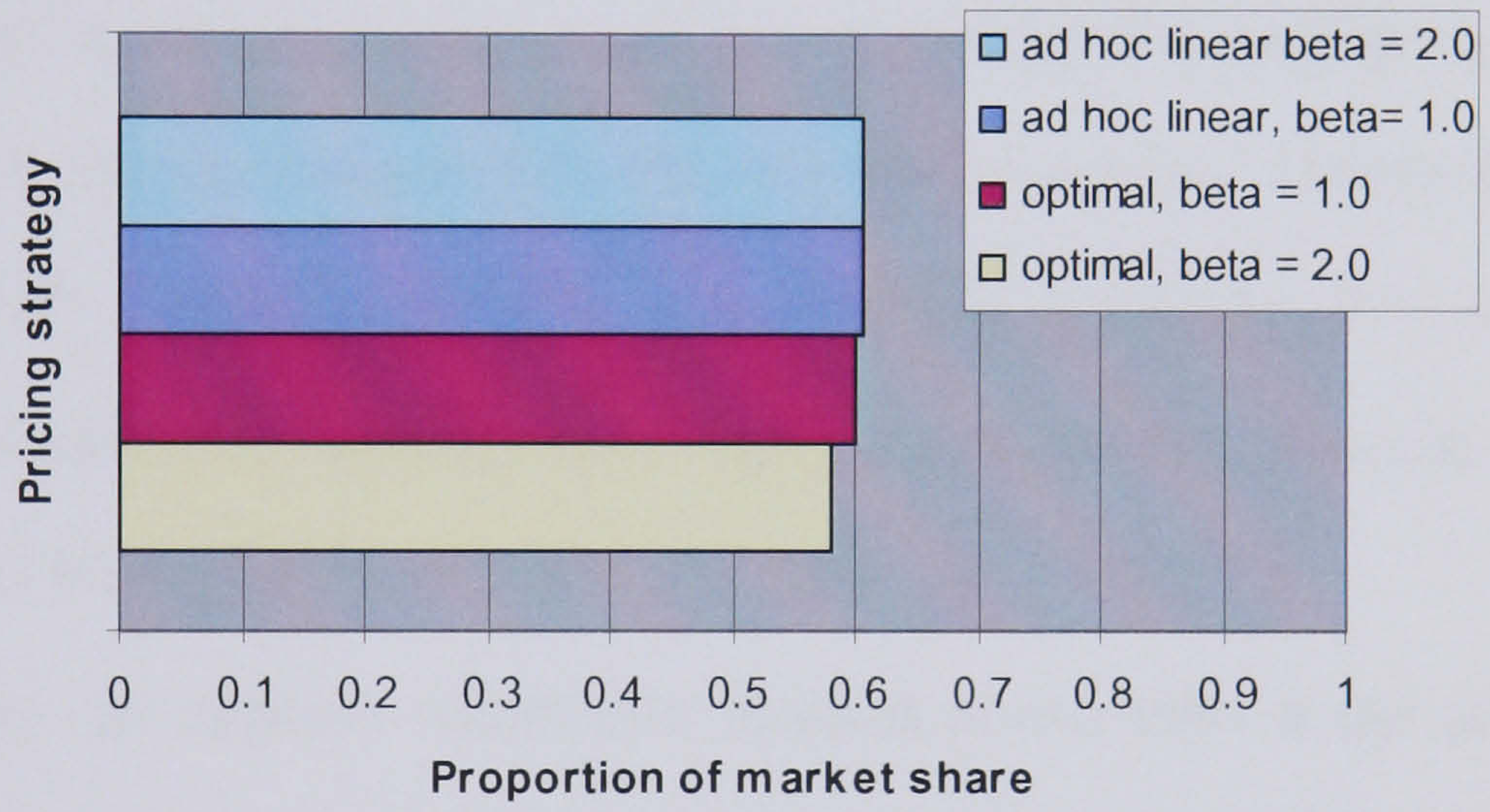

Figure 7-5 *Market share comparison of ad hoc and optimal pricing policies*

The predicted market share with the optimal dynamic pricing function is also similar to the proportion of users expected with the linear revenue attainment pricing function with low or medium preference for revenue attainment ($\varepsilon = 0.1$ or $\varepsilon = 0.5$) and inelastic demand ($\beta = 1.0$) (see Figure 7-6). In this case users would be attracted by relatively lower prices at the cost of higher call blocking. However, the optimal dynamic pricing strategy offers the best option by attracting a similar market share without relying on successful promotion or compromising the QoS of the service provided.

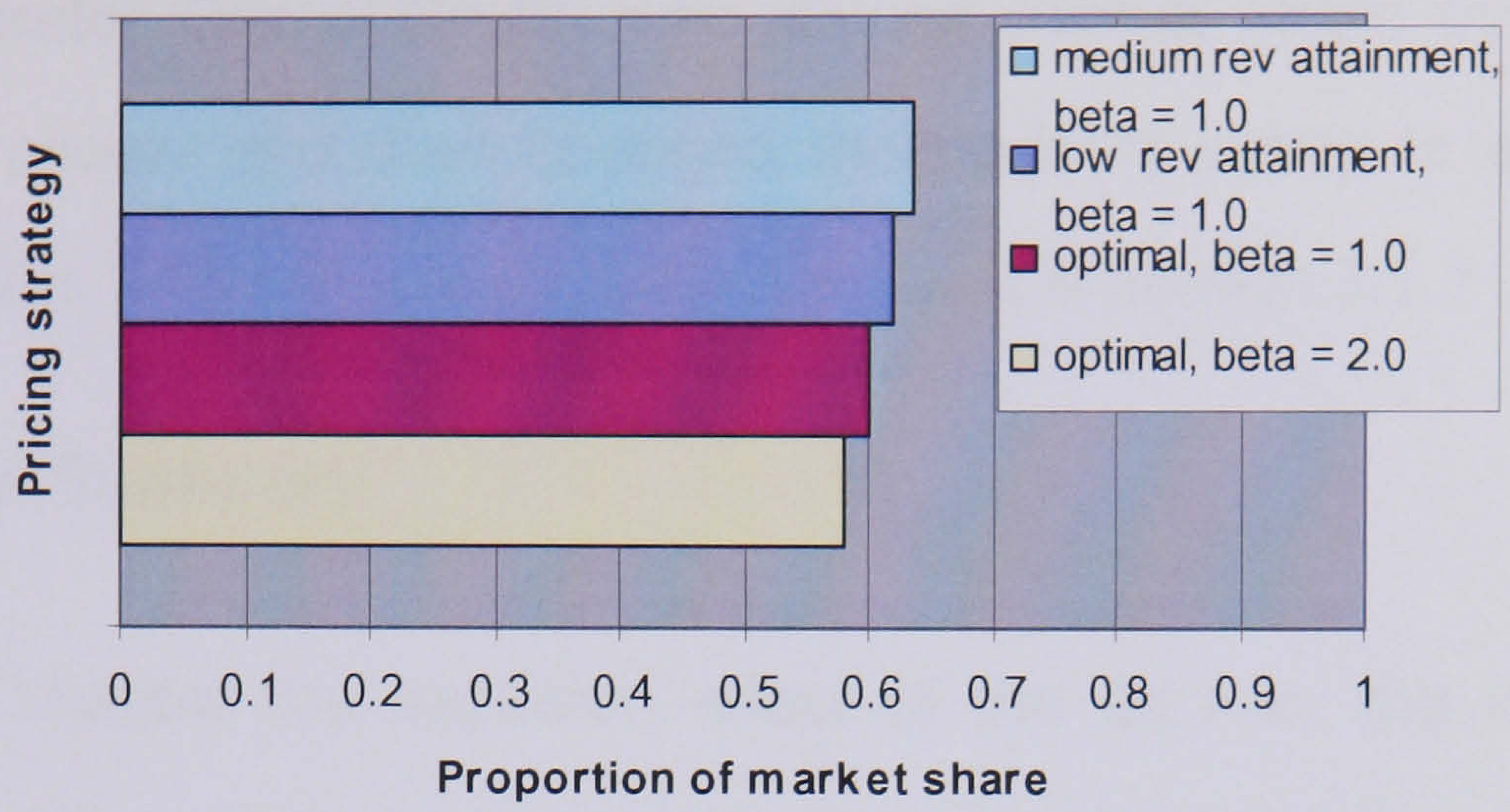

Figure 7-6 *Market share comparison of revenue attainment and optimal pricing policies*

The optimal dynamic pricing strategy would lead to a higher proportion of expected market share compared to the non-linear *ad hoc* competition driven pricing policy with inelastic demand (*i.e.* demand elasticity $\beta = 1.0$) and the linear revenue attainment strategy with high preference for revenue attainment ($\varepsilon = 0.9$, demand elasticity $\beta = 1.0$ and $\beta = 2.0$).

In summary, to capture maximum market share with a dynamic pricing policy the network operator should offer pricing strategies that lead to very low average prices. Those include the non-linear *ad hoc* competition driven pricing or the revenue attainment pricing with low preference for revenue attainment ($\varepsilon = 0.1$) and high preference for capacity utilisation ($\phi = 0.9$). However, these strategies will reduce the total revenue generated by the network and will increase the probability for call blocking, thus reducing the QoS provided. Therefore, in the long run the network operator has to consider pricing strategies that improve the QoS and generate the necessary revenue such as the linear *ad hoc* competition driven dynamic pricing strategy or the optimal dynamic pricing strategy. This is illustrated by the experience of the UK network operator One 2 One[52], who gained market share by offering very competitive prices, and then changed their pricing policy to ensure market share retention. This issue is discussed in detail is section 3.1.1.

7.2 Chapter Summary.

In this chapter the expected effect of the *ad hoc*, the linear revenue attainment and optimal revenue attainment pricing strategies on network operators' market share was predicted with the aim of identifying the optimal pricing for maximisation of the total user database. The results showed that dynamic pricing strategies could provide an increase in the number of subscribers for the network operator, compared to non-dynamically priced networks. The most marked increase in market share was with the dynamic pricing strategies offering lowest average prices. However, these strategies also lead to a significant increase in the probability of call blocking and retention of customers in a climate of low QoS would be unlikely.

The most attractive pricing strategy from both operator's and user's point of view is the optimal revenue attainment function, which generates the desired revenue for the network operator, while at the same time controlling call blocking. It also leads to significantly lower average prices in the network compared to the non-dynamic pricing and the competition driven *ad hoc* linear pricing strategies, which promotes an increase in the number of calls generated. The optimal dynamic pricing strategy offers an overall market share of 60% relative to a non-dynamically priced network[53] and these results

[52] T-Mobile since April 2002.

[53] This result is based on the assumption that the users have no objections to accepting a dynamic pricing policy.

support the hypothesis that dynamic pricing could be used successfully as a tool that serves the interests of both users and the network operator.

Chapter 8

8.1 Conclusions.

The aim of this project was to test the efficiency of dynamic pricing as a mechanism for load management and control in cellular mobile networks. The study involved the utilisation of tools and methods from a variety of disciplines (economics, engineering and control theory) which increased the complexity of the task significantly. After identifying the main parameters that affect the capacity of cellular wireless systems (GSM, GPRS and UMTS) and current suggested strategies for optimising network utilisation, an alternative "soft edge" approach for achieving better capacity utilisation through *dynamic pricing* (variation of price as a function of load) was suggested. The engineering and economic implications of the introduction of dynamic pricing were identified and possible dynamic pricing algorithms based on the load in the system suggested for both circuit-switched and packed-switched networks. The most challenging issue with dynamic pricing was identified as the mapping of system load to monetary units and determining the overall shape of the pricing function.

A mathematical model specifically developed for investigating the expected effect of dynamic pricing on the operator market share was presented. To model the expected effect of dynamic pricing on user behaviour a comprehensive mathematical model was developed taking into account the price elasticity of user demand as well as the existing pricing bias in the network. Substitution and time of day effects due to the fixed network were also modelled. In addition, a user mobility model showing the effect of price on user mobility was introduced by adapting a transportation model discussed by Wilson [78].

Three novel strategies for setting the dynamic prices as a function of system load were developed and tested using a seven-cell simulation model built in OPNET™. These were competition driven *ad hoc* pricing, linear revenue attainment pricing and optimal revenue attainment pricing (in which the requirement for linearity was dropped). Simulation results using the competition driven *ad hoc* dynamic pricing confirmed that the shape of the pricing function can significantly affect the behaviour of the network. In addition, the results identified generated revenue and number of blocked calls as the most sensitive factors in the system. Therefore, revenue attainment and control of the probability of call blocking were the two controlled variables considered in the control system model used for the development of the linear and optimal revenue attainment dynamic pricing strategies.

Results from the simulations showed that the revenue attainment pricing strategy was not a very efficient control function, as it failed to generate the target revenue, although it controlled call blocking. The optimal dynamic pricing strategy, on the other hand, generated the target revenue as well as controlling the percentage of blocked calls. In addition, the optimal dynamic pricing strategy led to a reduction in the average price in the network, promoting an increase in the number of calls, which makes it attractive from user's point of view. Therefore, this was the pricing strategy recommended to network operators.

Finally, the effectiveness of the different dynamic pricing policies in increasing the network operator's market share was tested. The results showed that the policies offering the lowest average prices would attract the largest proportion (up to 80%) of the market share. However, these strategies also led to very significant increases in the probability of call blocking and would, therefore, decrease the QoS that users receive. The optimal revenue attainment strategy would attract up to 60% of the market share with lower weighted average prices, while generating sufficient revenue. It has to be noted that the same proportion of the market would be captured with the linear *ad hoc* competition driven pricing, which generates significantly more revenue than the network without dynamic pricing by using higher prices. However, despite the higher prices it is likely that it could still attract users because the enhanced revenue means the network operator can invest more in promotions and advertisement. The results highlighted the difficulties with which the network operators are faced when choosing their pricing policy, as their market share is affected by a combination of factors whose influence can cancel each other out in the long run.

Dynamic pricing is a powerful tool which, when used correctly, can offer significant advantages to both the network operator and users of the cellular network. However, as this thesis showed, its effects on network performance are complex and further research is deemed necessary before practical implementation is attempted.

8.2 Future Directions.

Throughout this thesis a number of factors that affect the effectiveness of any dynamic pricing policy were identified as potential candidates for future research. Suggested areas for research that would enhance the understanding of the overall effect of dynamic pricing on all aspects of the cellular network management would be:

1. Extending the optimal dynamic pricing model to enable determination of optimal pricing for a network with a mix of voice and data type of traffic, taking into account the different traffic profile of voice and data.
2. Improving the mathematical model for user demand by taking into account the derivative of the price.
3. Modelling the effect of price on the length of the calls.
4. Further investigation into the optimal price update intervals, the practical implications of dynamic pricing on the signalling overhead in the network and the best method for conveying billing information to users.
5. Determination of the actual relationship between mobility elasticity α and price quasi-elasticity β together with an accurate estimation of these parameters from real time data.
6. Investigation of the optimal method for measuring the expected load in the network for accurate calculation of revenue attainment dynamic prices.
7. Further study and development of a discrete optimal revenue attainment dynamic pricing model.
8. Modelling the effect of the fixed line network on dynamic pricing.

Appendix A

System Parameters	North America (AMPS)	UK (TACS)	Japan (NTT)
Transmission Frequency (MHz) • Base station • Mobile station	870 - 890 825 - 845	935 – 960 890 - 915	870 – 885 925 - 940
Spacing between channels (kHz)	30	25	25
Number of channels	666/832	1000	600
Coverage radius by one base station (Km)	2 - 25	2 - 20	5 (Urban) 20 (Suburban)
Data transmission rate (kbs)	10	8	0.3

Table A-1 *First Generation Cellular Networks*

System Parameters	USA (IS-54)	Europe (GSM)	USA (IS-95)
Access technology	TDMA/FDMA	TDMA/FDMA	CDMA/FDMA
Transmission Frequency (MHz) • Base station • Mobile station	869 - 894 824 - 849	935 – 960 890 - 915	869 – 894 824 - 849
Spacing between channels (kHz)	30	200	1250
Number of channels per Radio Frequency channel	3	8	-
Data transmission rate (kbs)	7.95	13	8 (variable)

Table A-2 *Second Generation Cellular Systems*

Appendix B – UK Cellular Telecommunications Market.

B-1. Market Supply Structure - Network Operators and Service Providers.

In the UK market, the network operators are responsible for the planning and installation of the network, while the service providers are responsible for the distribution of the mobile handsets, the sale of subscriptions and the after-sale service. This structure encourages competition and the operators and service providers subsidise demand by selling mobile terminals at low prices and splitting the cost (see Figure B-1). Competition between operators and economies of scale in the market for hand held terminals are the key factors of the success of the British market.

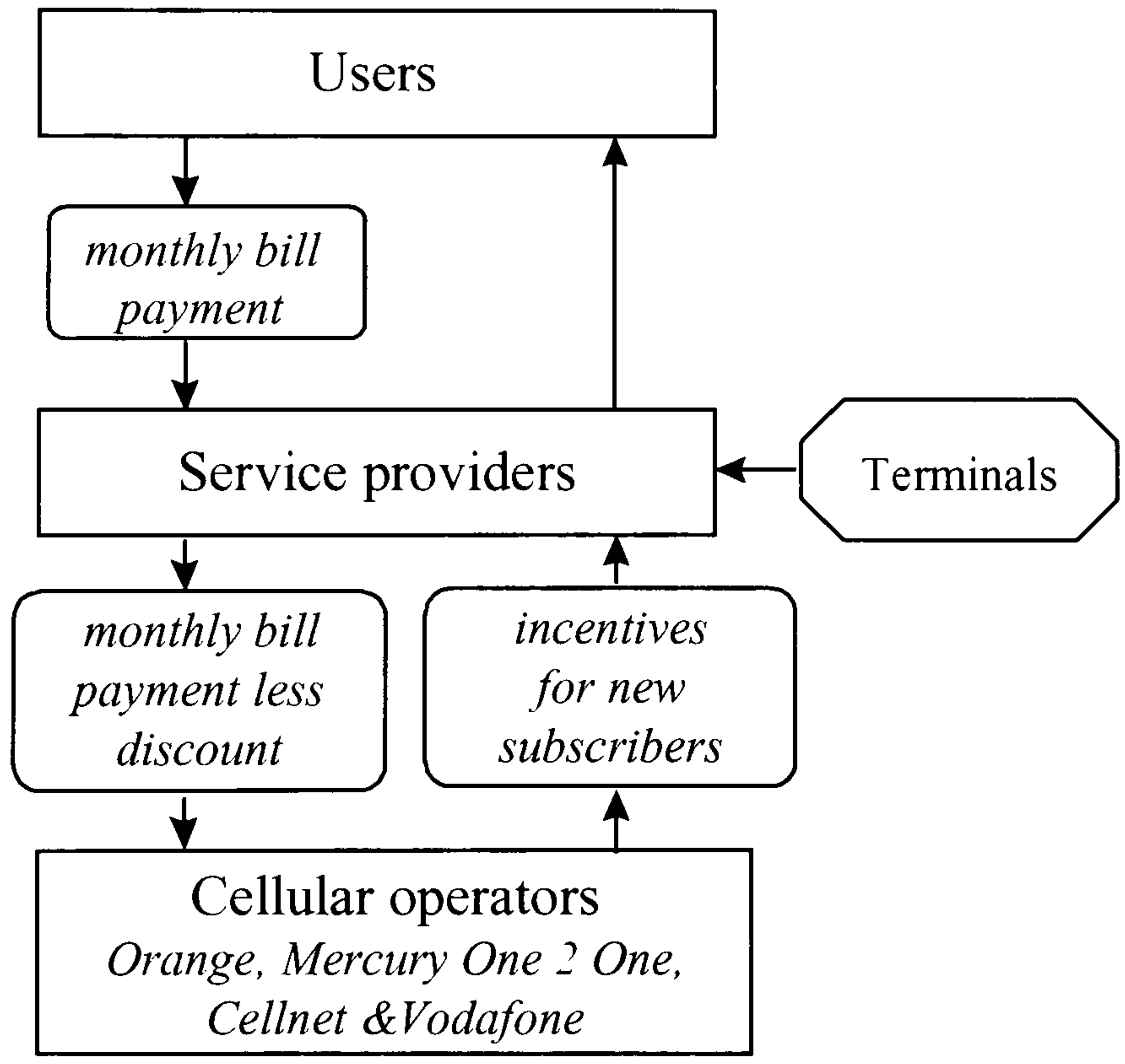

Figure B-1 *Structure of UK cellular phone distribution network*

There are four licenses granted to UK network operators, two for each 2nd generation digital standard. DCS1800 is used by T-Mobile[54] and Orange, and GSM is used by Cellnet and Vodafone. T-Mobile and Orange are relatively new entrants into the market, having launched their services in September 1993 and April 1994 respectively. Shown in Figure B-2 is the effect on price for peak and off-peak cellular calls at the launch of the additional competitors in the market [11].

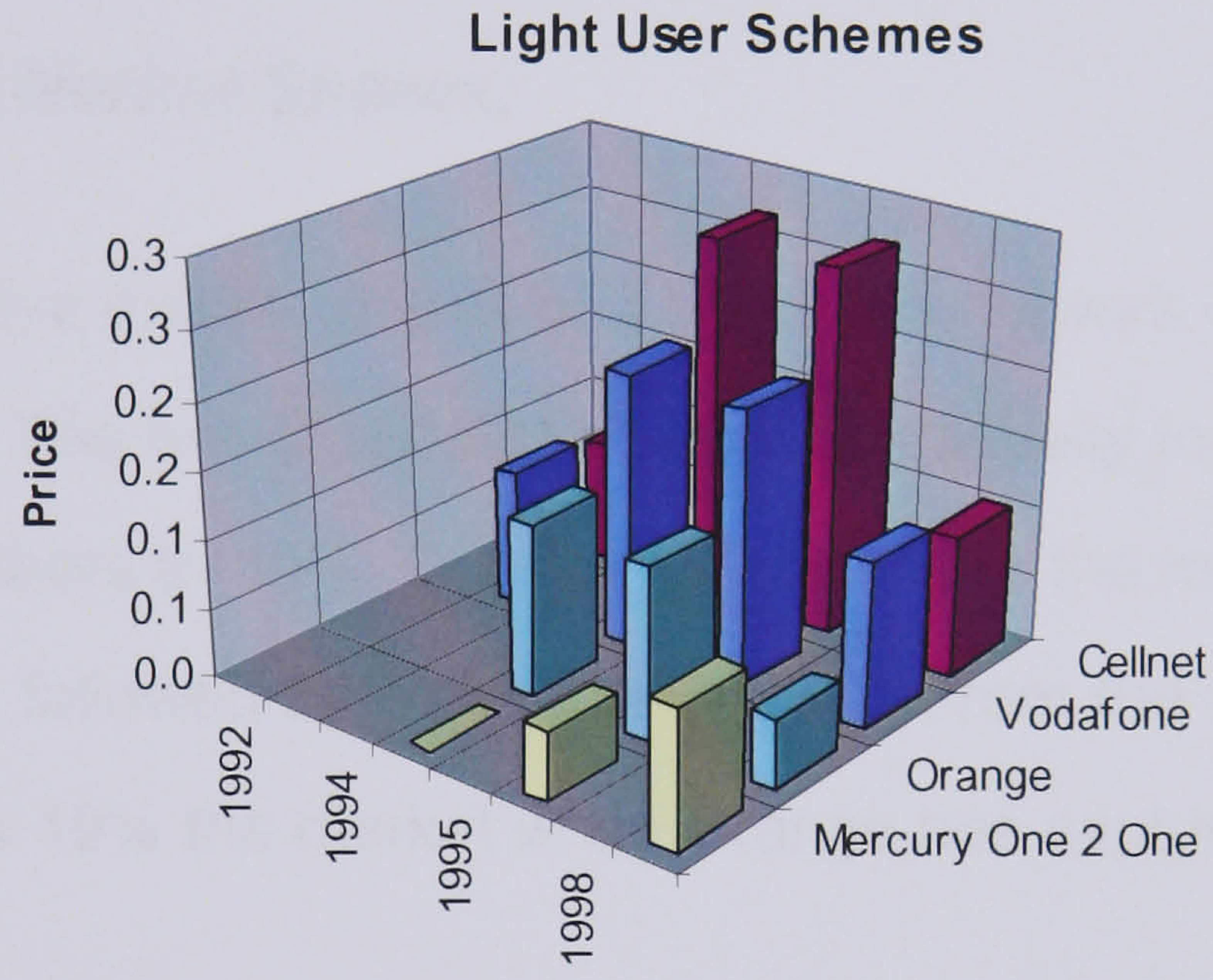

Figure B-2 *Prices of peak-hour calls from mobile phones* [55]

The most significant reduction in the tariffs of the older network operators Cellnet and Vodafone occurred when new competitors T-Mobile and Orange entered the market. This resulted in a 57% decrease in the peak-time tariffs of Cellnet and a 25% decrease in Vodafone's tariffs. This clearly demonstrates that direct competition drives prices down. In fact, after the initial introductory period, both One to One and Orange increased their prices by 16% and 13%

[54] [54] New name for One 2 One since April 2002.
[55] Direct comparison between tariffs is impossible due to the different amounts of free time included (bundled) in each tariff.

respectively, so that presently the prices charged by all the operators are quite similar.

Licences for 3rd generation networks were distributed by auction and there will be five operators – British Telecom, Vodafone, Orange, One2One and Telesystem International Wireless Inc. (TWI) from Canada. The increase in the number of competitors in the market will lead to active competition between the service operators in the future as they attempt to capture and retain their market share.

B-2 Relative Market Shares.

The relative market shares of the four UK network operators can be seen in Figure B-3. The two DCS1800 operators gradually increased their share of mobile subscribers to 40%. Currently Vodafone is the market leader with 32% of the market, followed by Cellnet with 27%. From the late entrants T-Mobile currently holds 19% the market while Orange has captured 22% of the market [109].

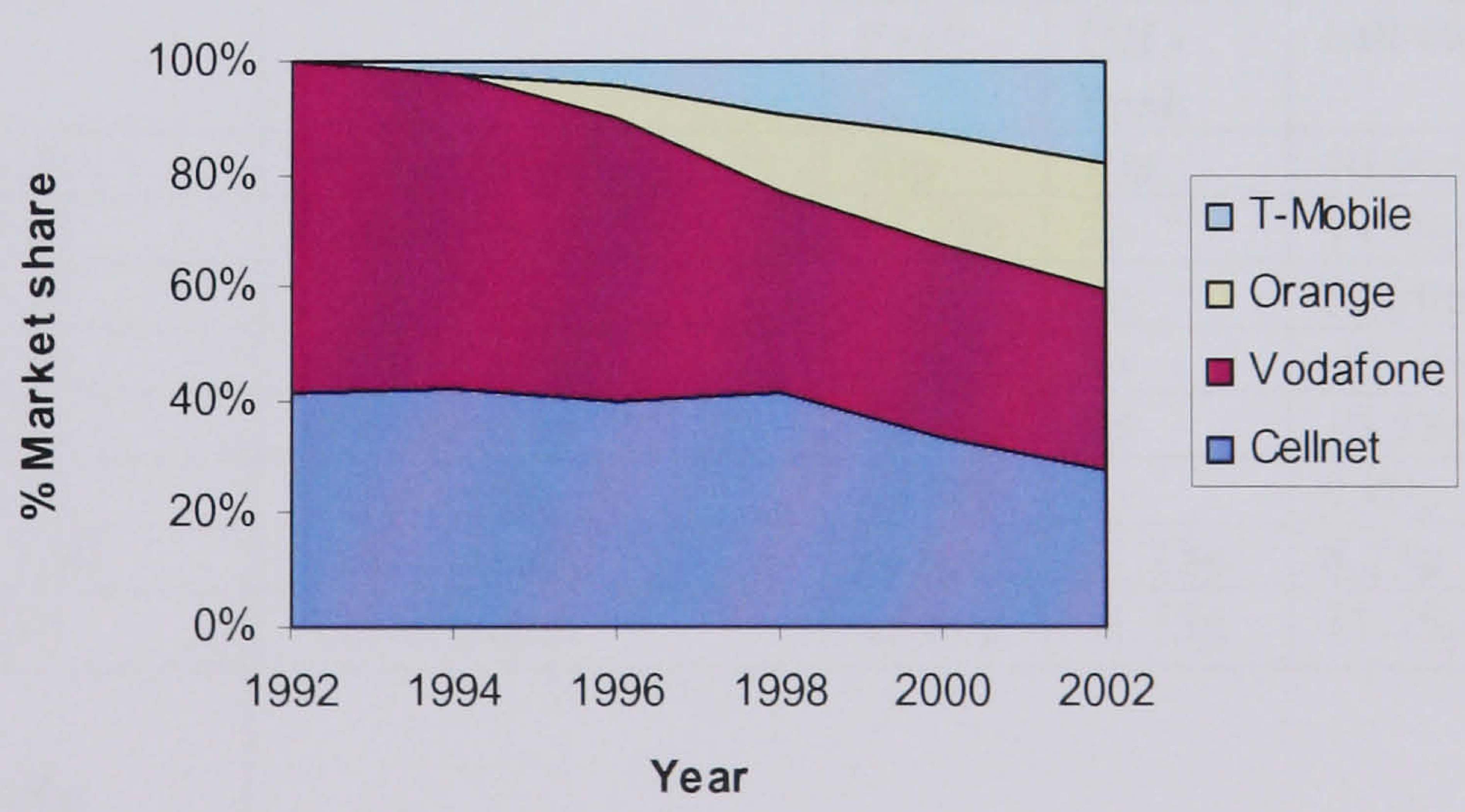

Figure B-3 *Market shares of the four UK network operators.*

B-3. Cellular Phone Tariffs as on 30th March 1999.

Tariff	**Pre-Pay**	**Occasional Caller+**	**Regular Caller+**	**Frequent Caller+**
Connection	included	£ 29.79	£ 29.79	£ 35.25
Monthly Line Rental	included	£ 14.89	£ 21.28	£ 34.04
Inclusive Calls	£ 4.25	£ 4.25	£ 12.77	£ 17.02
Peak Call Charges	34.04p	34.04p	29.79p	20.00p
Off - Peak Charges	10.21p	10.21p	10.21p	10.21p
Charges	10.21p	10.21p	10.21p	10.21p

Table B-1 *Cellnet tariffs*

Tariff	**Minute Saver 5**	**Minute Saver 20**	**Minute Saver 60**	**Minute Saver 100**
Connection	35.25	35.25	35.25	35.25
Monthly Rental	16.99	19.99	26.99	37.49
Charge Period	8am-8pm Mon-Fri	8am-8pm Mon-Sat	8am-8pm Mon-Sat	8am-8pm Mon-Sat
Charge/min	40p	35p	32p	21p
Charge Period	All other times	All other times	All other times	All other times
Charge/min	10p	10p	10p	10p
Inclusive Air time	5 minutes	20 minutes	60 minutes	100 minutes

Table B-1 *Vodafone tariffs.*

Plan Name	**Monthly Charge**	**Talk Time included (per month)**	**Call Charges**		**Orange to Orange peak call charges**
			Peak	**Off - Peak**	
everyday 20	£15.00	20 min off-peak daily	40p	10p	40.00p
talk 15	£17.63	15 minutes	29.38p	5p	14.69p
talk 30	£17.50	30 minutes	30p	5p	30.00p
talk 60	£29.38	120 minutes	23.50p	5p	11.75p
talk 200	£58.75	400 minutes	21.15p	5p	10.58p
talk 360	£88.12	720 minutes	18.80p	5p	9.40p
talk 540	£117.50	1080 minutes	16.45p	11.75p	8.23p
Talkshare+60	£44.06	120 minutes	23.50p	11.75p	11.75p

Table B-2 *Orange tariffs.*

Time Plan	**Monthly Charge**	**Free Time**	**Peak Calls**	**Off-Peak Calls**	**VoiceMail**
One-2-Evening	£22.50	Local only, Mon-Fri, 6pm-midnight	30p	10p	free

One-2-Weekend	£17.50	Local only , all weekend	30p	10p	free
One-2-45	£17.50	45 mins Anytime	30p	5p	free
Precept 100	£25.00	100 mins Anytime	20p	5p	free
Precept 200	£40.00	200 mins Anytime	15p	5p	free
Precept Daytime	£59.00	Weekday daytime, local calls	15p	10p	free

Table B-1 *One2One tariffs.*

Appendix C – Survey Methodology and Findings.

To investigate the attitudes and preferences of mobile telephone users a survey, in the form of a questionnaire, was conducted in the spring of 1998. Series of questions enquired about the consumers' current behaviour and the degree of their satisfaction or dissatisfaction with the currently offered service. A different set of questions was asked to establish the preferences of the customers. Some questions tested the consumer's preference for "per-use" or "bulk" pricing. Other questions examined users' concept for Quality of Service in relation to clarity of voice and connectivity of the speech. Finally, the users were asked to speculate about the future possible uses of mobile phones. Probing in depth about the opinions and preferences of the mobile phone users would allow us to draw conclusions about their behaviour.

C-1. Methodology.

A questionnaire (see end of appendix) was published on the Internet in March 1998 and various Internet users were invited to complete it. To minimise the professional bias in the respondents, the URL address of the questionnaire was sent by e-mail to two specific user groups - students and staff at University College London and employees at British Telecom Labs, Ipswich (private and business users). In total there were 93 responses to the survey, with 29 from a company email addresses.

The results were analysed using Statistical Product and Service Solutions for Windows (SPSS) [110]. Two tests were applied to assess if a significant proportion of the test population have given the same answers. The first test, Chi-square, was used to test for significance when studying the population as a whole. The second test, One-way Anova, was applied when the test population was divided into multiple categories and significant differences were sought between the categories. For example, to test if users would accept dynamic pricing, a Chi-square test would be applied to the results while to determine if there is a difference in the answers of business and private users a One-way Anova test would be used for analysis.

Despite the limited number of people who responded to the survey, the results from this study can be used as an indication of mobile phone users' opinions and attitudes of and can be used as a basis for assumptions and further studies.

C-2. Results and Discussion.

In designing the survey, it was ensured that three main areas of interest were covered by the questionnaire - the sensitivity of users to price and QoS, as well as their attitudes to dynamic pricing. Overlapping questions were asked to guard against question bias.

C-2.1. User attitudes to pricing and QoS.

To determine the relative importance of price for mobile users, a question enquired about the main factors determining the choice of mobile operator (see Figure C-1). The majority of respondents (34.4%) indicated that price of the calls was the deciding factor in their choice. The second most significant factor (at 25%) was coverage area. Although, about 5% of business user chose a subscriber because they could get a package at a special discounted price, the main incentive for their choice is still the price. There is no significant difference in the preferences expressed by the two groups (business and private), which can perhaps be explained by the fact that the majority of users pay for their calls themselves rather than have the cost covered by a company.

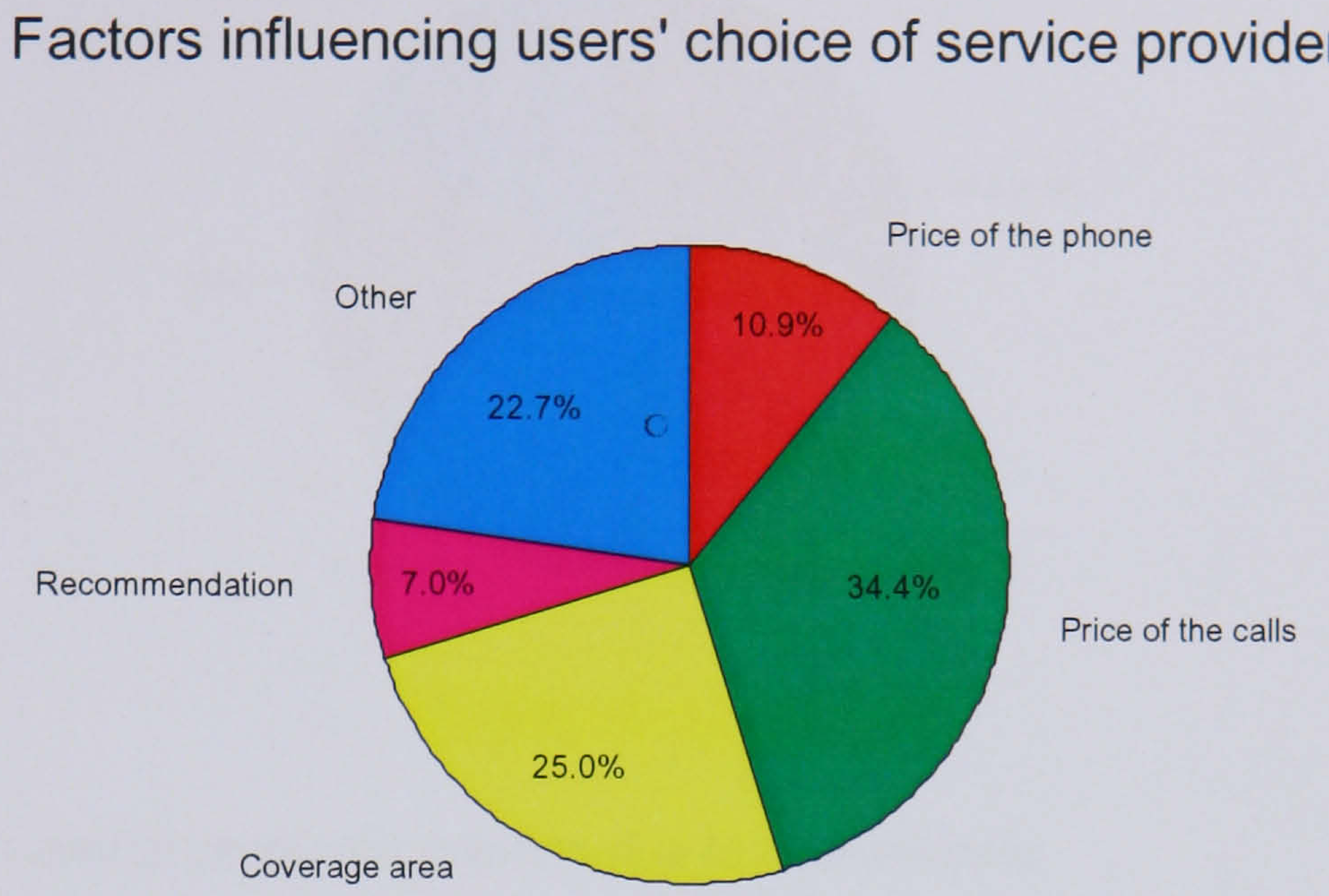

Figure C-1 *Main factors influencing the choice of service provider*

Therefore, users will choose the service provider with lowest call charges and adequate coverage area. Price is an important factor in determining user choices and if the introduction of dynamic pricing were to reduce the price of calls to and from mobile phones, it would give a competitive advantage to the respective operator.

The second factor that could influence users' choice of service provider is QoS provided by the network. First, users' reactions to call dropping and call blocking were assessed. Surprisingly, call blocking seemed to present no significant problem for the users with 42% of users not finding it at all frustrating and only 17% getting very annoyed by it (see Figure C-2).

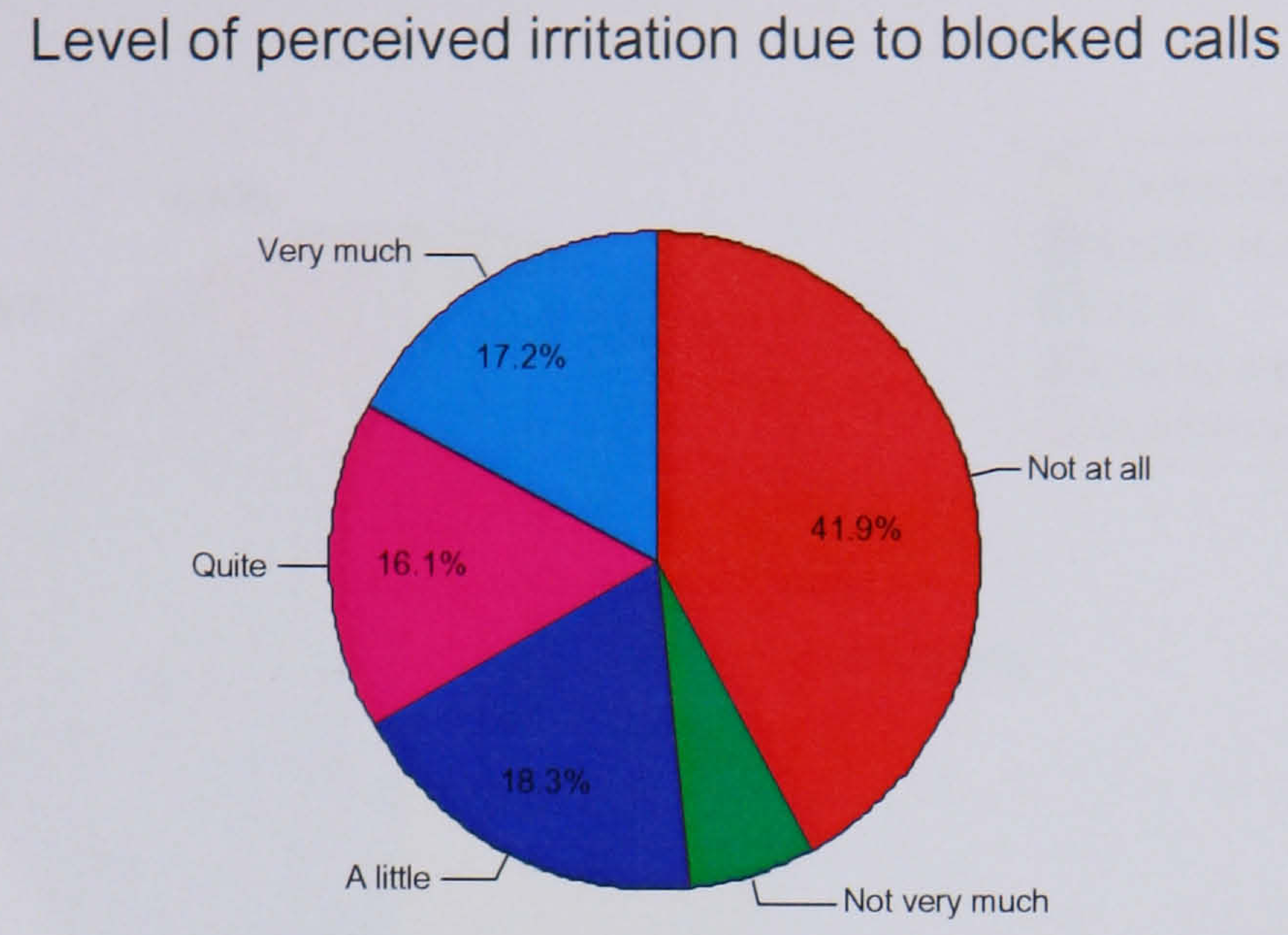

Figure C-2 *Level of perceived irritation due to blocked calls*

However, when asked about their reason for purchasing the mobile telephone 74% indicated that contactability (the ability to make and receive calls) and emergency use have been their main consideration (see Figure C-3). This contradicts the finding that the call blocking is not a serious problem. A wider survey and a more targeted questioning may resolve this ambiguity. Therefore, the ability to make and receive calls is paramount to users perception of acceptable QoS and care has to be taken to ensure that access to the network is guaranteed most of the time.

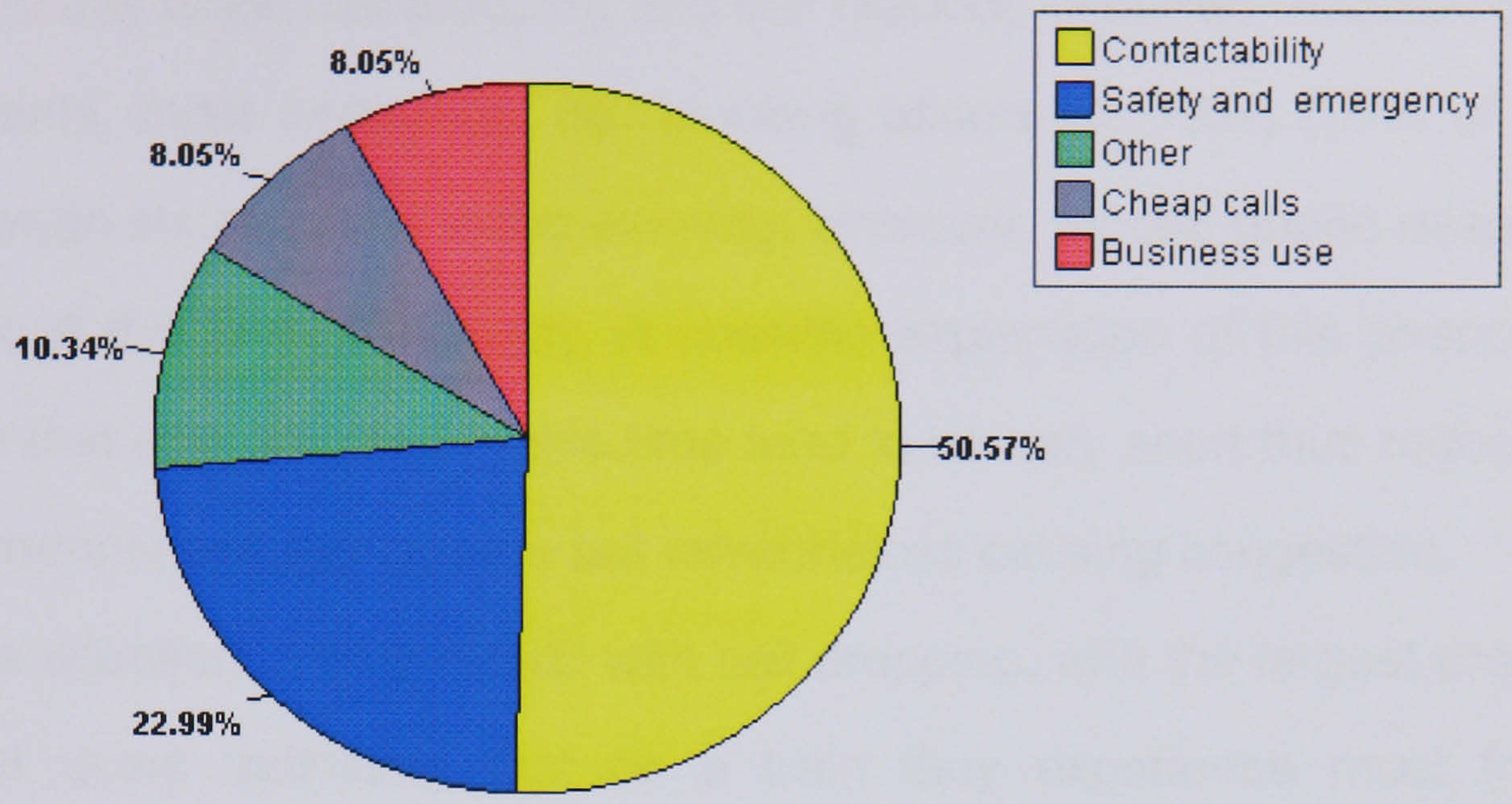

Figure C-3 *Reasons for purchasing the mobile phone*

In contrast to call blocking, a significant proportion of users (34.4%) find call dropping very irritating (see Figure C-4).

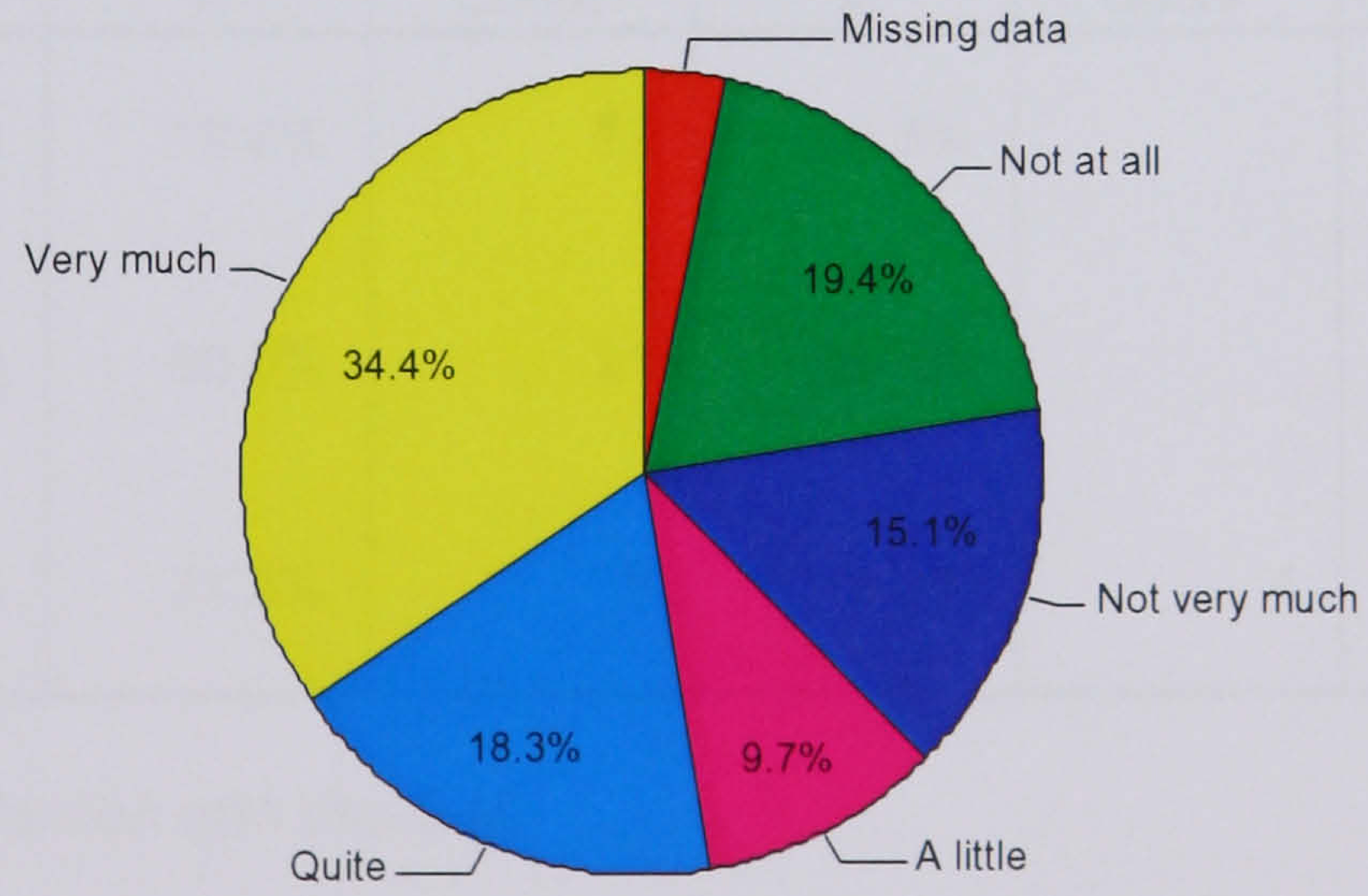

Figure C-4 *Level of perceived irritation due to dropped call*

A set of questions tried to ascertain the most common places and times during the day when call dropping and call blocking occurred. According to the respondents, most frequently, call blocking occurs on trains (35% of cases) and between six and nine in the evening. However, mobile phone usage does not peak at this time of the day. A possible explanation of this phenomenon could be that calls initiated at this time tend to be very short thus reducing the overall channel occupancy time but nevertheless causing congestion.

The situation is very similar with call dropping, with the largest proportion (29%) of users indicating that on a train they experience most frequent interruptions. This can perhaps be accounted for by the fact that trains often travel at high speed, crossing cells frequently and requesting many handoffs, which causes the problem. Overall, there is a significant difference in the perception of QoS provided by the cellular networks for the two groups (Table C-1 shows the frequency distribution of the replies).

Satisfaction with QoS

	User group					
	Private users		Business users		Missing data	
	QoS		QoS		QoS	
	Count	%	Count	%	Count	%
Not very satisfied	9	18.8%	12	29.3%		
Fairly satisfied	24	50.0%	21	51.2%		
Very satisfied	15	31.3%	8	19.5%	4	100.0%

Table C-1 *Overall satisfaction with the QoS*

29.3% of the business users from BT are not satisfied with the service, compared to only 18.8% of the private users from UCL. This perhaps can be explained by the higher expectations business users have for mobile phone services.

When asked if they are prepared to pay more for a better QoS a significant proportion of the users said no (see Table C-2). We can interpret this in two ways: either users are happy with their current service or they do not think that QoS is as important as price. However, there is a significant difference in satisfaction with QoS between Orange and One to One users, which indicates that people are aware of the QoS provided by the network. Therefore, although users are not prepared to pay more for better QoS, they are sensitive to different qualities of service a network provides.

Readiness to pay more for better service

	User group					
	Private users		Business users		Missing data	
	Pay for better service		Pay for better service		Pay for better service	
	Count	%	Count	%	Count	%
Yes	6	12.5%	6	14.6%	2	50.0%
No	32	66.7%	24	58.5%	1	25.0%
Don't know	10	20.8%	11	26.8%	1	25.0%

Table C-2 *Customers readiness to pay more for better service*

In summary, the results from the survey suggest that the price of the calls is the most important factor in determining users' choice of service provider. Users are aware of the QoS provided by the network and are very irritated by call dropping; therefore the service operators can increase user satisfaction by implementing preventative measures such as channel allocation algorithms prioritising channel availability for handover, for example.

C-2.2. User attitudes to dynamic pricing.

The second factor tested by the survey is the expected change in user behaviour at the introduction dynamic pricing. Intuitively, one would expect users to reduce their usage of the network when prices are high and increase it when prices are lower. To check the viability of this hypothesis, users' preferences for tariff types and behaviour were questioned. Firstly, users preferences for "bulk" and "pay as you use" tariffs were determined. A significant proportion of the test population indicated that they would prefer the "pay as you use" tariffs (see Table C-3).

Preference for per-use rate tariffs

	User group					
	Private users		Business users		Missing data	
	Prefer flat rate tariffs		Prefer flat rate tariffs		Prefer flat rate tariffs	
	Count	%	Count	%	Count	%
Yes	14	29.2%	11	26.8%	2	50.0%
No	33	68.8%	19	46.3%	2	50.0%
Don't know	1	2.1%	11	26.8%		

Table C-3 *User preference for "per-use" tariffs*

This is in contrast to finding by Cosgrove and Linhart [78] for the USA market. This difference can possibly be explained with the different prevailing cultures in USA and Europe. Traditionally, in USA the telephone companies have used fixed tariffs for local calls, whereas in UK the "pay as you use" tariffs have been more popular. As a result the customers have adjusted their preferences accordingly and users in USA prefer the fixed rate tariffs. This is confirmed by the finding that from the 93 respondents. The proportion of users who pay for their calls have a significant preference for "per use" payment, whereas the proportion of users who do not pay for their calls, have expressed no particular preference (see Table C-4).

Preference for per-use rate tariffs by users paying for their calls

	Paying for calls			
	Yes		No	
	Prefer flat rate tariffs		Prefer flat rate tariffs	
	Count	%	Count	%
Yes	25	30.9%	2	16.7%
No	50	61.7%	4	33.3%
Don't know	6	7.4%	6	50.0%

Table C-4 *Preference for per use tariffs by users who pay for their calls*

The results from the survey also indicate that a significant proportion of users would find it helpful to be reminded the price of the calls before they make a call as it would help them plan their usage better (see Table C-5)

Perceived usefulness of seeing price before making call

	User group					
	Private users		Business users		Missing data	
	Price will help plan better		Price will help plan better		Price will help plan better	
	Count	%	Count	%	Count	%
Yes	19	39.6%	22	53.7%	4	100.0%
No	13	27.1%	8	19.5%		
Depending on the call	15	31.3%	10	24.4%		
Don't know	1	2.1%	1	2.4%		

Table C-5 *Perceived usefulness of seeing price before making a call*

In fact, a significant number of respondents have implied that they will use their phone more if prices were variable but lower than current nominal prices (see Table C-6). These findings are particularly important for justifying the implementation of dynamic pricing

Change in use depending on price

	User group					
	Private users		Business users		Missing data	
	Use more if price is lower		Use more if price is lower		Use more if price is lower	
	Count	%	Count	%	Count	%
Yes	30	62.5%	25	61.0%	3	75.0%
No	11	22.9%	9	22.0%	1	25.0%
Don't know	7	14.6%	7	17.1%		

Table C-6 *Usage if prices are variable but lower*

Most important, however, is the confirmation of the repression phenomena observed by Cosgrove and Linhart [78]. A significant proportion of users have indicated that they are conscious of reducing the duration and frequency of their calls during peak hours. Surprisingly, this is true even for people who do not pay for their calls (see Table C-7). This finding indicates that potentially dynamic pricing could lead to shift in usage from busy hours with high prices to non-busy hours with lower prices.

Change in use depending on time of day

	User group					
	Private users		Business users		Missing data	
	Keep calls shorter during peak hours		Keep calls shorter during peak hours		Keep calls shorter during peak hours	
	Count	%	Count	%	Count	%
Yes	32	66.7%	17	41.5%	4	100.0%
No	3	6.3%	9	22.0%		
Depending on the call	13	27.1%	15	36.6%		

Table C-7 *Usage reduction during peak hours*

To summarise, the results from the survey indicate that price plays a major role in user's decision making process and that users would favour a pricing policy that would reduce overall call charges. Users have also stated that they reduce their usage of the phone during the more expensive peak hours, verifying the hypothesis that higher prices would discourage usage. Although, neither the business nor the private users are prepared to pay more for higher QoS, both groups have indicated that they are particularly irritated by call *dropping*, which suggests that load management techniques can be employed to increase user satisfaction with the performance of the cellular network.

C-3. Survey Questions.

1. For how long have you been using a mobile phone?
    Less than 1 month
    1-6months
    6months-1 year
    1-2years
    More than 2years

2. Why did you purchase the phone?

3. Which is your service provider
    Cellnet
    One2One
    Vodafone
    Orange

4. Do you have a digital or analogue system
    Digital
    Analogue
    Don't Know

5. What influenced your choice of service provider
    Price of the phone
    Price of the calls
    Coverage area
    Recommendation
    Other (please specify)

6. Compared to your expectations do you think that your mobile phone service is:

Reliable: Doesn't cut off
Allows access at all times
Other (please specify)

The quality of voice is: Clear
Continuous
Noisy
Noisy but intelligible

Other (please specify)

7. How often do you see on your display the message "Limited Service Available" which only allows you to make emergency calls
    - More than twice per day
    - 1-2 times per day
    - 2-3 times per week
    - About once per week
    - Less than once per week
8. How annoying do you find this where '1' = 'Not at all', '5'= 'Most Annoying

9. When and where does this usually happen
    - Before 9am
    - Between 9am 12pm
    - Between 12pm 3pm
    - Between 3pm 6pm
    - Between 6pm 9pm
    - After 9pm

    - On the motorways
    - On the train
    - In the suburbs
    - In the centre of town
    - Other places

10. How often do your calls get cut off in the middle of a conversation
    - More than twice per day
    - 1-2 times per day
    - 2-3 times per week
    - About once per week
    - Less than once per week

11. How annoying do you find this where '1' = 'Not at all', '5'= 'Most Annoying

12. When and where does this usually happen
    - Before 9am
    - Between 9am 12pm
    - Between 12pm 3pm
    - Between 3pm 6pm
    - Between 6pm 9pm
    - After 9pm

    - On the motorways
    - On the train
    - In the suburbs
    - In the centre of town
    - Other places

13. Do you pay for your calls?
    - Yes
    - No

14. How much do you pay for your calls on the mobile
    - Peak:
    - Off-peak:

Don't know

15. When making a call during peak hours do you keep your calls shorter to reduce the total cost?

    Yes
    No
    Depends on the call

16. If the prices of the calls are displayed before you start your call would it help you to plan and decide the call duration?

    Yes
    No
    Depends on the call
    Don't know

17. Would you like the option of paying more for a call if this meant receiving better service

    Yes
    No
    Depends on the call
    For what in particular would you pay?

18. If the prices were variable (changing by time and location), but less than the current rate, would you use your mobile more often?

    Yes
    No
    Don't know

19. Would you prefer a pre-fixed payment for all your calls (local and national) at the beginning of the month as opposed to 'per call' charge?

    Yes
    No
    Don't know
    Please add any additional comments

20. Do you have or foresee in the future a need for laptop wireless services (using your mobile phone like a computer

    Yes
    No

Appendix D

D-1. Matlab Routine for Finding the Balancing Factors A_i and B_j.

% ELENASOLVER FOR FINDING THE BALANCING FACTORS – this is the master function called ELENASOLVER.m, which calls the slave function ELENAFUNCT.m.

```
% v is the vector whose elements are x1, x2, y1 and y2
% It will be the solution to the minimisation problem

% Start with an arbitrary choice of values, all ones
% Edit this line if other starting values are desired
x1 = 70; x2 = 50; y1 = 0.02; y2 = 0.01;

vo = [x1 x2 y1 y2];

% set up some parameters for the solver
% It retuns a solution when the error returned by Elenafunct.m
% is below Tolfun = 0.0001 and the values in v are steady from
% one iteration to another to within TolX = 0.0001
% (good  choices are problem dependent).
% (Type help optimset to discover more about parameters)

options = optimset('MaxFunEvals',8000,'MaxIter',8000,...
  'TolFun',1e-4,'TolX',1e-4);

% Send the initial guess into the solver. The equations are
% in the Matlab program called elenafunct.m

v = fminsearch('elenafunct',vo,options);
abs(v)  % This line prints the result
ELENAFUNCT.m is the 'slave' optimisation function called by the
'master' program called ELENASOLVER.m
function [error] = elenafunct(v);
% ELENAFUNCT is the 'slave' optimisation function called by the
% 'master' program called ELENASOLVER.m
% v is the vector of variables that are being optimised.
% Force a real non-negative solution:
v = abs(v); % Negative values not permitted. It works like
% this: even if the optimiser master routine searches negative values
% we only ever use the positive value. This trick is needed
% because the MATLAB optimiser is unconstrained, but we want
% to exclude negative values.
% Unpack vector v
x1 = v(1); x2 = v(2); y1 = v(3); y2 = v(4);
% calculate residuals
e1 = x1 - 1/(0.6*y1+0.1*y2);
e2 = x2 - 1/(0.8*y1+0.2*y2);
e3 = y1 - 1/(0.9*x1+0.2*x2);
e4 = y2 - 1/(1.2*x1+0.4*x2);

% Sum of squares.
```

```
error = e1^2+e2^2+e3^2+e4^2;

% To see the detailed working, delete the % sign on
% to get a run time print out of all the trials.
% error = e1^2+e2^2+e3^2+e4^2

% The variable "error" is now returned to the MATLAB
% fsolvemin programme which will try some other x1, x2, y1, y2
```

D-2. Matlab Routine for Finding Competition Driven Dynamic Prices.

```
%The program follows the notation developed in the thesis.
pmin = 0.01;
pmax = 0.5;
m = 1.2*sqrt(1+(pmax-pmin)^2);
z = sqrt(m^2 - (0.5-0.01)^2);
test = inline('1.2432-(sinh(x))/(x)');
x = fzero(test, 2);
c = 1/(2*x);
h = (atanh((pmax-pmin)/m)*2*c-1)/2;
lambda1 = c*cosh(h/c)-pmin;
%lambda2 = (c*cosh((1+h)/c)-pmax);
%price1 = c*cosh((q.+h)/c)-lambda1;
price2 = c*cosh((q/100+h)/c)-lambda1;
```

D-3. Matlab Routine for Finding Revenue Attainment Dynamic Prices.

```
%This routine searches and find the optimal ψ that will allow the attainment of a given
%revenue. "test.m" is the slave function called to minimise the difference between the actual
%and desired revenue.
global Rcur, b, a, q, pint, C, tau;

bigcount = 1;
Ropt_arr = [22, 11, 7, 5, 4, 6, 15, 50, 397,519,520,516,517,509, 575, 615, 612, 540,
125, 85, 82, 76, 68,38];
a_arr = [31, 15, 8, 5, 6, 7,22, 86, 152, 237, 239, 231,224, 216, 262, 314, 325, 239, 244,
160, 149, 130, 117, 68];
xo = 0.05;
b = 1.0;
ax = [0.1:0.1:1];
C = 80;
tau = 1.2;
people = (31*7);

for bigcount= 1:1:25,
```

```
  if (( bigcount > 9) & (bigcount<=19))
     pint = 0.25;
  else
     pint = 0.05;
  end

  Ropt = Ropt_arr(bigcount);
  a = (7*a_arr(bigcount));
  counter = 1;
  Rcum = 0;
  n = 0;
  ave = 0;

 Rcum = 0;
    for t = 1:1,
       Rcur = (Ropt-Rcum)/(13-t);
       [x1, xval] = fminsearch('test',xo);
       psy(t,counter)= x1;
       Rcum = Rcum + (a*exp(-b*((x1*q+0.01)-pint))*(x1*q+0.01)*tau);
       deltap(t,counter)=(x1*q+0.01)-pint;
       q = (a*exp(-b*((x1*q+0.01)-pint)))/C;
       pr(t,counter)=(x1*q+0.01);
       Att(t, counter) = Rcur;
       Cum(t, counter)= Rcum;
    end
  counter = counter+1;
  n = n+1;
  bigcount= bigcount+1;
end
dlmwrite('test1',test_rev,'\t');
dlmwrite('test2',test_a,'\t');
dlmwrite('test3',test_psy,'\t');
dlmwrite('test4',test_att_rev,'\t');
dlmwrite('test5',test_pr,'\t');

%"test.m"
function [error] = test(x);
global Rcur, a, q, pint, C, tau, mew;

m = x(1);
if((((m*q)<0)|(-a*exp(-b*((m*q+0.01)-pint))+C)<0)&(mew>0))
   e1=rand(1)*8000;
else
   e1 = (Rcur-(a*exp(-b*((m*q+0.01)-pint))*(m*q+0.01)*tau))- mew*log(m*q)- mew*log(-
a*exp(-b*((m*q+0.01)-pint))+C);
end
error = e1.^2;
```

D-4. C Routine for Finding Revenue Attainment Dynamic Prices.

```
#include<math.h>
#include<stdio.h>
#define MAXIT 5000
```

```
#define FACTOR 1.1
#define NTRY 500

float rtsafe(int a, double rev, double alfa, double beta, double pr_b, double c_len, double
n, float *result)

/* Using a combination of Newton Raphson and bisection, find the root */
/* of a function bracketed between x1 and x2.  The root, returned as the  */
/* function value rtsafe, will be refined until its accuracy is known within */
/* +/- xacc.  funcd is a user-supplied routine that returns both the function */
/* value and the first derivative of the function.                  */

{
  void funcd(int a, double rev, double alfa, double beta, double pr_b, double c_len,
double n, float x, float *fun ,float *der);
  int zbrac(int a, double rev, double alfa, double beta, double pr_b, double c_len, double
n,float *x1, float *x2);
  int j;
  float df, dx, dxold, f, fh, fl;
  float temp, xh, xl, rts, t1, t2;
  float x1, x2;

  x2 = 1/(alfa+beta);
  x1 = x2-0.02;
  zbrac(a, rev, alfa, beta, pr_b, c_len, n, &x1, &x2);

  funcd(a, rev, alfa ,beta, pr_b, c_len, n, x1, &fl, &df);
  funcd(a, rev, alfa ,beta, pr_b, c_len, n, x2, &fh, &df);
  /* //printf("min = %f, max = %f \n", fl, fh); */

  if ((fl>0.0 && fh>0.0) || (fl< 0.0 && fh< 0.0))
          printf("Root must be bracketed in rtsafe");
  if (fl == 0.0) return x1;
  if (fh == 0.0) return x2;
  if (fl< 0.0){
          xl = x1;
          xh = x2;
  } else {
          xh = x1;
          xl = x2;
  }
  rts = 0.5*(x1+x2);
  /* //printf("RTS1 %f\n", rts); */
  dxold = fabs(x2-x1);
  dx = dxold;
  funcd(a, rev, alfa,beta, pr_b, c_len, n, rts, &f, &df);
  for (j = 1; j<=MAXIT; j++){
          if ((((rts-xh)*df-f)*((rts-xl)*df-f) > 0.0) || (fabs(2.0*f)>fabs(dxold*df))){
          dxold = dx;
          dx = 0.5*(xh-xl);
          rts = xl+dx;

          if (xl == rts)
                  {
                  *result = rts;
```

```
                        return rts;
                        }
        }
        else {
                dxold = dx;
                dx = f/df;
                temp = rts;
                rts -= dx;
                /* //printf("RTS3 %f\n", rts); */
                if (temp == rts)
                {
                *result = rts;
                return rts;
                }
        }

        if (fabs(dx) < 0.0001) {*result = rts; return rts;}
        funcd(a, rev, alfa, beta, pr_b, c_len, n, rts, &f, &df);
        if (f<0.0)
                fl = rts;
        else
                xh = rts;
        }
        printf("Maximum number of iterrations exceeded in rtfsafe \n");
        return 0.0;

    }

    void funcd(int a, double rev, double alfa, double beta, double pr_b, double c_len, double
n, float x, float *fun ,float *der)
    {
        double sorted;
        double people;

        if (x < 0.05)
        {
                sorted = 1.0/(10*(x-pr_b));
                *fun = rev - (a*exp((-beta)*(x-pr_b)) + pow(sorted, beta))*x*c_len;
                *der = (beta*a*exp((-beta)*(x-pr_b))+ 10*beta*pow(sorted, beta+1))*x*c_len -
(a*exp((-beta)*(x-pr_b)) + pow(sorted, beta))*c_len;
                /*//printf("Function low is %f\n", *fun);
                //printf("People low are %f\n", people);*/

        }
        else
        {
                *fun = rev - a*exp((-beta)*(x - pr_b))*x*c_len;
                /* //people = 7*a*a*n*exp((-alfa-beta)*(x - pr_b));*/
                *der = beta*a*exp((-beta)*(x - pr_b))*x*c_len - a*exp((-beta)*(x - pr_b))*c_len;
                /* //printf("Function high is %f\n", *fun);
                //printf("People high are %f\n", people); */
        }

    }
```

```
int zbrac(int a, double rev, double alfa, double beta, double pr_b, double c_len, double n,float *x1, float *x2)
{
  int j;
  float f1, f2, der;
  void funcd(int a, double rev, double alfa, double beta, double pr_b, double c_len, double n, float x, float *fun ,float *der);

  funcd(a, rev, alfa, beta, pr_b, c_len, n, *x1, &f1 ,&der);
  funcd(a, rev, alfa, beta, pr_b, c_len, n, *x2, &f2 ,&der);

  for (j=1; j<=NTRY; j++) {
  if (f1*f2 < 0.0) return 1;
  if (fabs(f1) < fabs(f2))
        funcd(a, rev, alfa, beta, pr_b, c_len, n, *x1+=FACTOR*(*x1-*x2), &f1 ,&der);
  else
        funcd(a, rev, alfa, beta, pr_b, c_len, n, *x2+=FACTOR*(*x2-*x1), &f2 ,&der);
  }
  printf("Did not converge\n");

  return 0;
}
```

D-5. Matlab Routine for Finding Optimal Dynamic Prices.

```
%This routine calls function "opt.m" which finds the optimal prices in the network by calling %"opt_2.m" which finds the revenue attainment price, and then chooses the optimal of those %in terms of optimal load in the network. The solutions are found using the numerical %method of exhaustive enumeration.

global number
for number = 1:1:24,
   [price(number), der(number), error(number), quan(number)] = feval('opt',number);
end
wk1write('opt_price1_10_0', price);
wk1write('opt_der1_0', der);
wk1write('opt_quan1_0',quan);

%"opt.m" with N denoting the derivative of the optimal dynamic price.
function [optimal_price, derivative, first_one, q] = opt(bigcount);
global Rcur, b, a, q, pint, N, p, q0;

%bigcount = 1;
i = 1;
Ropt_arr = [20, 14, 7, 5, 4, 7, 19, 65, 524,768,790,791,774,751, 828, 902, 912, 786, 198, 122, 117, 105, 95,54];
a_arr = [31, 15, 8, 5, 6, 7,22, 86, 152, 237, 239, 231,224, 216, 262, 314, 325, 239, 244, 160, 149, 130, 117, 68];
b = 1;
q = 0.2;
q0 = 0.0;
i = 1;
```

```
j = 1;
k = 1;
first_one = 2*80^2;
other= 1;
%for bigcount= 1:1:1,

if (bigcount == 1)
   q0 = 0;
else
   q0 = 7*a_arr(bigcount-1)/12;
end
   if (( bigcount > 8) & (bigcount<19))
      pint = 0.25;
      xo = 0.40;
   else
      pint = 0.05;
      xo = 0.40;
   end
   %The second equation is:
   cor = 0;
   spec = 1;
   for N= 0.01:0.01:1/b,
      %or_pr(k) = N;
      %or_pr(special)= N;
      %h1 = -(0.01*cor);
      h1 = 0;
      %h2 = ((1/b-(0.01*cor))-0.01);
      h2 = 0;
      j = 1;
      for p = h1:0.01:h2,
         %der_price(k,j)= p;
         %der_price(spec) = p;
         %or_pr(spec)= N;
         i = 1;
         %for q = 0.1:0.1:0.1,
         a = (7*a_arr(bigcount));
         q = a/20;
         Rcum = 0;
         Rcur = Ropt_arr(bigcount);
         [x1, xval] = fminsearch('opt_2',xo);
         %pr= x1;
         u(k,j)=((4799.9+1.44*N*x1)/(2-1.728*p*x1))*(1-2*exp(-0.6*p*x1))+2*exp(-
0.6*p*x1)*q0;

         if (((u(k,j)>0)&(u(k,j)<80)))
            diff_u(other)= u(k,j);
            der_price(other) = p;
            or_pr(other)= N;
            R(other)         =         ((1/(2-1.7*p*x1))^2)*(1.44*exp(-0.6*p*x1)*N*(2*(-2
+exp(0.6*p*x1))*q*(-2+1.7*p*x1)+(4000+1.2*N*x1)*(-5.77+exp(0.6*p*x1)*(4.88-1.7*p*x1))));

            if( N >= 0.05)
               dem(other) = a*exp(-b*(N-pint))/12;
            else
               if (b==1)
```

```
            numb = 10.0;
            %numb
          else
            numb = 10.0;
            %numb
          end

          dem(other) = a*exp(-b*(N-pint))+(numb*N)^(-b)/12;
          %other_demand = dem(other)
        end

        difference = (80 - dem(other))^2 +(diff_u(other) - dem(other))^2;
        if(difference < first_one)
          first_one = difference;
          optimal_price(i) = N;
          derivative(i) = p;
          %optimal_price(i)= N
          %derivative(i) = p
          %first_one
          i= i+1;
        end
        other = other+1;
      end
      %u(k,j)
      %load(k,j) = q;
      %price(k,j)= x1;
      i= i+1;
      %end
      j = j+1;
      spec = spec+1;
    end
    cor = cor+1;
    k = k+1;
  end
  %bigcount= bigcount+1;
  %end
plot3(or_pr,der_price,diff_u,'*:')
xlabel('Price')
ylabel('Derivative')
zlabel('Demand')
```

Appendix E - Detailed Solution of the Ad Hoc Competition Driven Pricing Model.

The optimisation problem is

$$Minimise \int_{Q_{min}}^{Q_{max}} (P_{min} + P(q))dq\,, \tag{E-1}$$

subject to

$$\sqrt{1+P(q)'^{2}}\; dq = M \tag{E-2}$$

and the boundary conditions:

$$P(Q_{min}) = P_{min} \tag{E-3}$$

$$P(Q_{max}) = P_{max} \tag{E-4}$$

P_{min} - Minimum price;

P_{max} - Maximum price;

q - Load in the network;

$P(q)$ - Intermediate prices in the network;

M - Competition market price;

Q_{min} - Minimum load to start dynamic pricing;

Q_{max} - Maximum load in the network.

Taking $Q_{min} = 0$ and using calculus of variations, the curve minimising the area bound by itself can be found using Lagrange's multiplier (see Butkov [98] for proof for sufficiency and necessity conditions). Thus we have to solve:

$$H(P, P', q) = \int_{0}^{Q\max} \left(P_{min} + P(q) + \lambda\sqrt{1 + P(q)'^{2}} \right) dq \tag{E-5}$$

As (5-5) does not explicitly dependent on q a special formula for finding the functional can be used (see Irving and Mullineux [99]) and the Euler equation which will be solved becomes:

$$P_{\min} + P(q) + \lambda\sqrt{1+P(q)'^2} - \lambda\frac{P(q)'^2}{\sqrt{1+P(q)'^2}} = const = K \qquad \text{(E-6)}$$

Rearranging equation (E-6) gives:

$$\lambda = (K - P_{\min} - P(q))\sqrt{1 + P(q)'^2} \qquad \text{(E-7)}$$

and expressing in terms of $P(q)'$ we get

$$P(q)' = \frac{\sqrt{\lambda^2 - (K - P_{\min} - P(q))^2}}{K - P_{\min} - P(q)} \qquad \text{(E-8)}$$

Let $K - P_{\min} - P(q) = \lambda\cos u$ and substituting into (E-8) and changing variables gives:

$$\frac{dP}{dq} = \frac{\lambda\sin u}{\lambda\cos u}\frac{du}{du} \qquad \text{(E-9)}$$

$$\frac{dP}{du} = \lambda\sin u = \frac{\lambda\sin u}{\lambda\cos u}\frac{dq}{du} \qquad \text{(E-10)}$$

and this gives

$$\frac{dq}{du} = \lambda\cos u \qquad \text{(E-11)}$$

$$q = -\lambda\sin u + h \qquad \text{(E-12)}$$

$$\sin u = \frac{h-q}{\lambda}. \tag{E-13}$$

Using right hand angle triangle equalities and changing the variable back to q gives:

$$\sqrt{\lambda^2 - (h-q)^2} = K - P_{\min} - P(q) \tag{E-14}$$

The general solution to this optimisation problem is:

$$P(q) = K - P_{\min} - \sqrt{\lambda^2 - (h-q)^2} \tag{E-15}$$

To solve the problem for boundary conditions ((E-3)) and (4) and determine an exact solution the following system of equations has to be solved:

$$\begin{aligned} P(0) &= P_{\min} = K - P_{\min} - \sqrt{\lambda^2 - h^2} \\ P(Q_{\max}) &= P_{\max} = K - P_{\min} - \sqrt{\lambda^2 - (h - Q_{\max})^2} \\ &\int_0^{Q_{\max}} \sqrt{1 + P(q)'^2}\, dq = M \end{aligned} \tag{E-16}$$

The integral is evaluated by differentiating (E-15) which gives:

$$P(q)' = \frac{h - q}{\sqrt{\lambda^2 - (h-q)^2}} \tag{E-17}$$

and setting $h - q = \lambda \cos\theta$

$$\begin{aligned} &\int_0^{Q_{\max}} \frac{\lambda}{\sqrt{\lambda^2 - (h-q)^2}} dq = \int_0^{Q_{\max}} \frac{\lambda}{\lambda \sin\theta} \frac{dq d\theta}{d\theta} = \\ &\int_0^{Q_{\max}} \lambda d\theta = \lambda\theta \Big|_0^{Q_{\max}} = \cos^{-1} \frac{(h - Q_{\max})}{\lambda} - \cos^{-1}\left(\frac{h}{\lambda}\right) = M \end{aligned} \tag{E-18}$$

The constants λ and h can be found by solving the following system of equations numerically.

$$P_{\max} - P_{\min} = \sqrt{\lambda^2 - h^2} - \sqrt{\lambda^2 - (h - Q_{\max})^2} \tag{E-19}$$

$$M = \cos^{-1}\frac{(h - Q_{\max})}{\lambda} - \cos^{-1}\left(\frac{h}{\lambda}\right)$$

Appendix F - Additional Graphs from the Simulations.

F.1 Simulation Model Verification.

F.1.1 Effect of Mobility Elasticity α on Call Blocking.

To validate the effect of user mobility on the performance of the network, the simulation was be run by changing the mobile elasticity (parameter α in the mathematical model) and keeping all other variable in the model constant. To facilitate correct operation of the mobility model, the prices charged in different cells were different and taken randomly from the interval $[0.05, 0.35]$ units but kept constant throughout the simulation. The simulations were run for a 24-hour period using the variables presented in Table 5-1.

As users move through the network their spatial distribution will affect call blocking in the system. User behaviour is determined by mobility elasticity α. When $\alpha < 0$ the users will be acting irrationally and move to more expensive cells and when $\alpha > 0$ move to cheaper cells. Therefore, it can be expected that the number of blocked calls will initially decrease as α increases and eventually increase again as the proportion of users shifting to cheaper cells increases and cell capacity begins to saturate.

Shown in Figure F-1 is the total number of blocked calls in the system as mobility elasticity α increases. The total number of blocked calls in the network initially decreases as α increase, but then begins to increase again after an inflection point at $\alpha = 2$. The inflection point at $\alpha = 2$ rather than $\alpha = 0$ is due to the exponential relationship in the user distribution model. The proportion of users moving to a particular cell depends on the mobility elasticity α and the price in the cell P_i. For $\alpha < 0$, the shift will be towards the cell with the $\max\{P_i\}$ in the system.

If $\alpha > 0$ the shift will be towards the $\min\{P_i\}$ in the system. The magnitude of the mobility exponential function for $\alpha < 0$ is $\alpha P_i > 1$, while if $\alpha > 0$, the magnitude of the function is $0 < \alpha P_j < 1$ and, therefore, the mobility with $\alpha < 0$ will generate a greater proportion of blocked calls than the mobility with $\alpha > 0$.

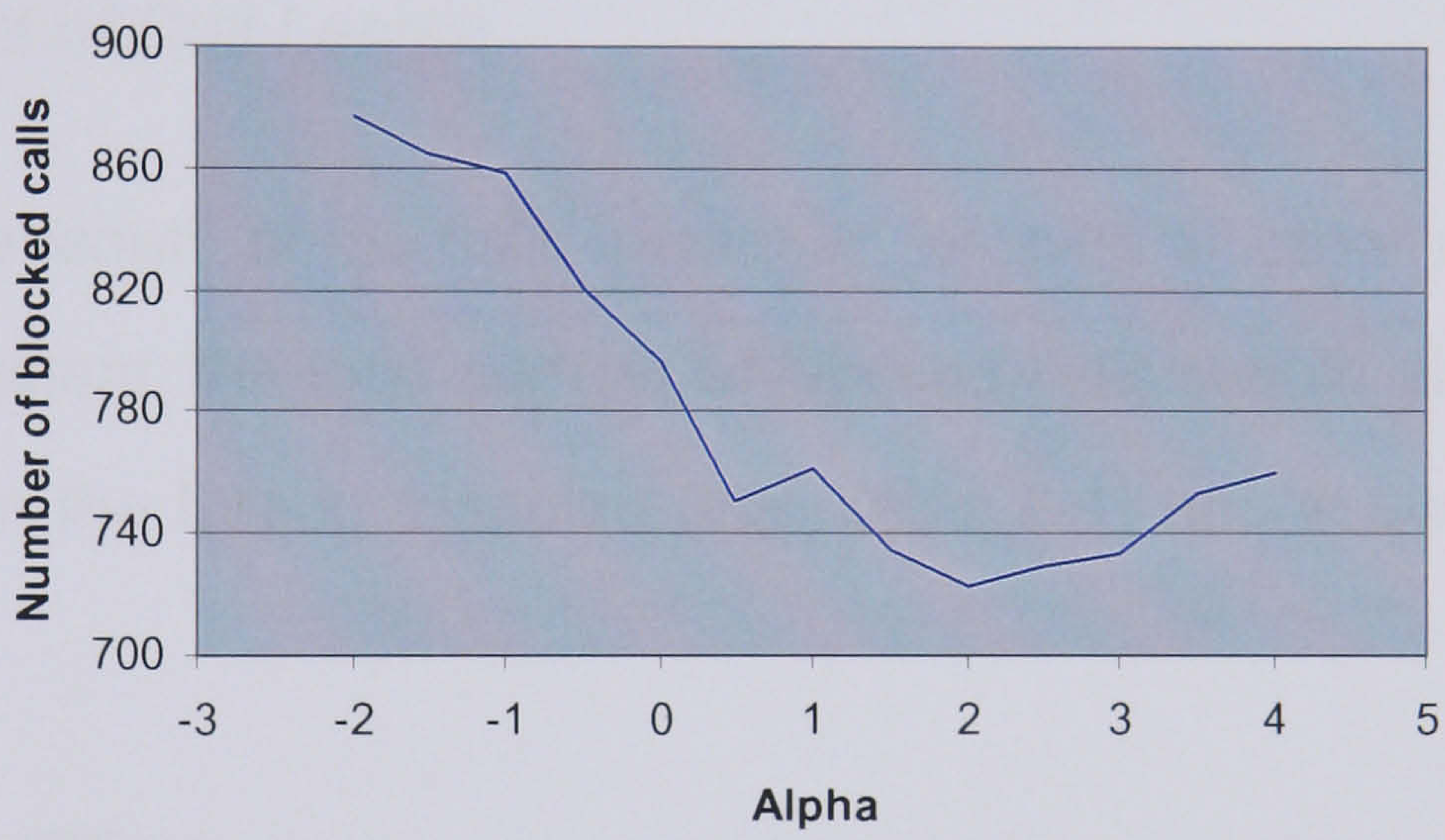

Figure F-1 *Effect of mobility elasticity α on call blocking*

F.1.2 Effect of demand elasticity β on Call Blocking.

The price elasticity of user demand is effected by the parameter β in the mathematical model developed in chapter 4. It represents the sensitivity of users to price changes. If the sensitivity of users to price increases (keeping the price constant) the total number of calls generated in the network at any given price will decrease. As a result the total number of blocked calls should also decrease and Figure F-2 shows that the simulation behaves as expected.

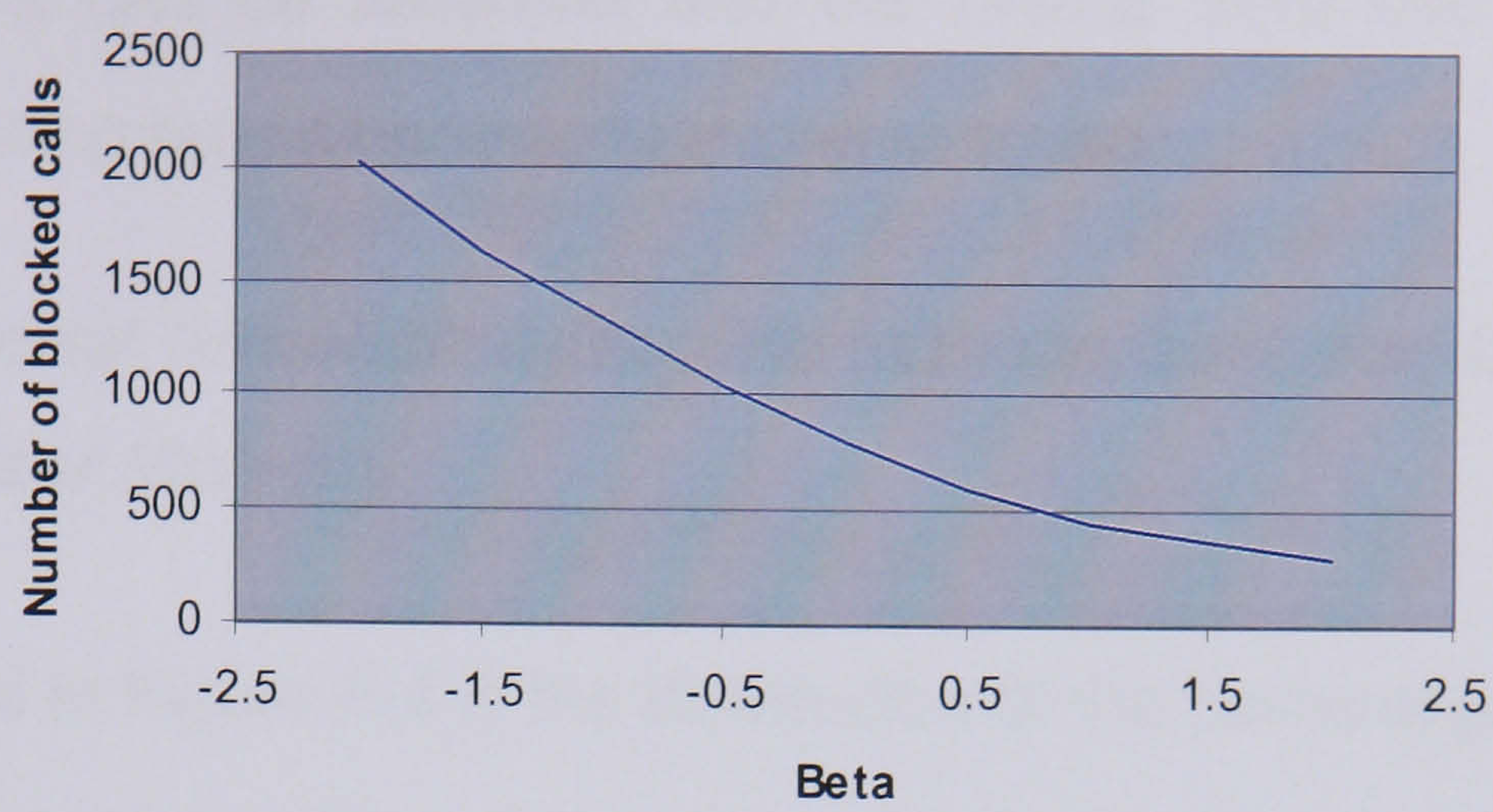

Figure F-2 *Effect of demand elasticity β on the number of blocked calls*

F.1.3 Effect of Call Length.

As the length of the calls increases, keeping all other parameters in the system constant, the total number of blocked calls should also increase. This follows from the Erlang formulae (see Table 2-1) and is confirmed in Figure F-3.

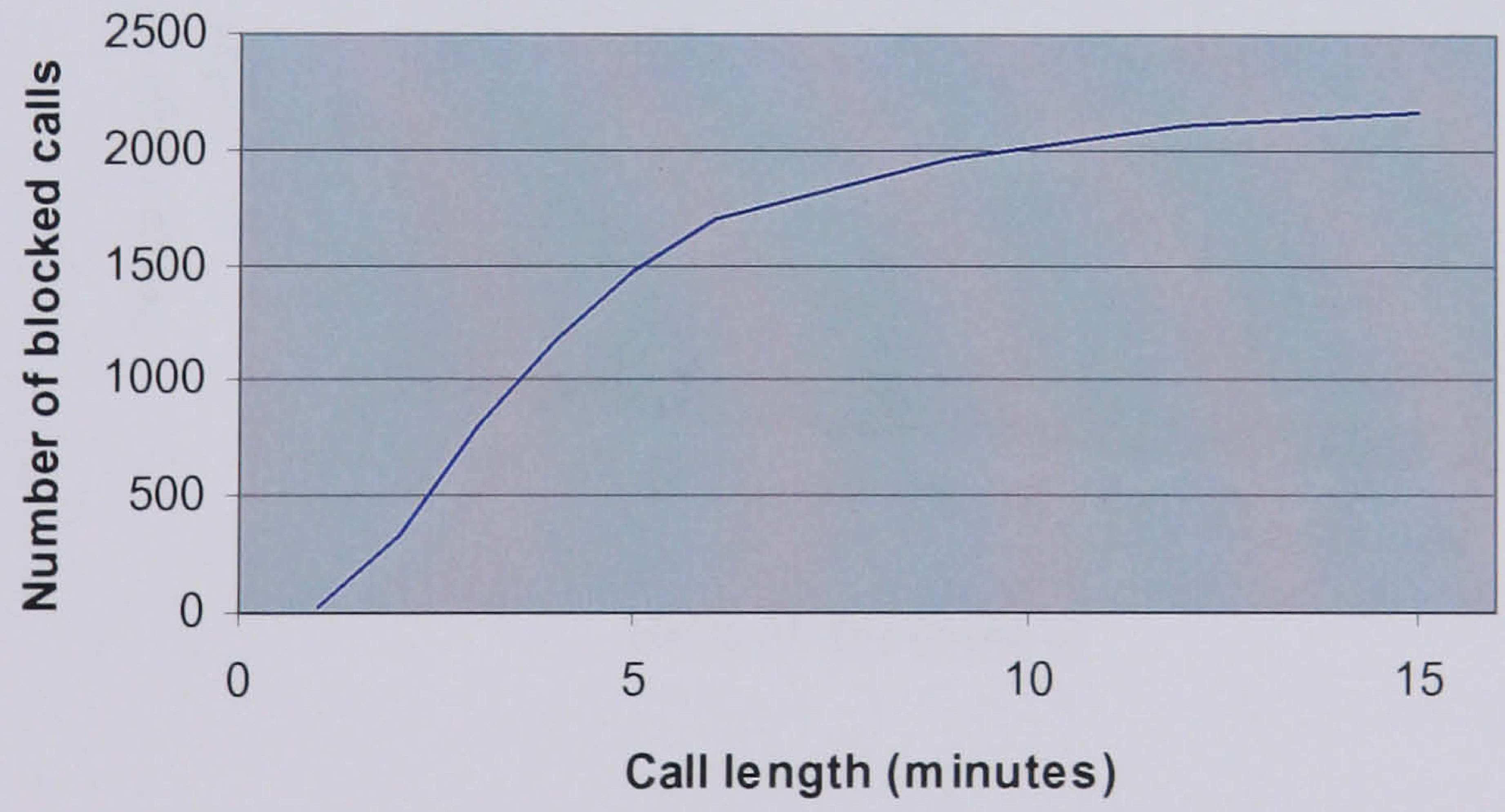

Figure F-3 *Effect of call length on the number of blocked calls*

These results indicate that the simulation system behaviour is consistent with the expected analytical behaviour in the case of non-dynamic prices.

Therefore, it can be assumed that the results from the system with the dynamic pricing operating would also be accurate.

F.2 Additional Simulation Results with Competition Driven Pricing without User Mobility.

Plotted in Figure F-4 is the distribution of the percentage of blocked calls as a function of the time of the day. As demand becomes unit elastic the majority of call blocking occurs during off-peak hours, rather than the peak-hours. This is due to the effect of the substitution effect with the fixed network, which leads to a very significant increase in the number of users attempting to access the network during off-peak hours.

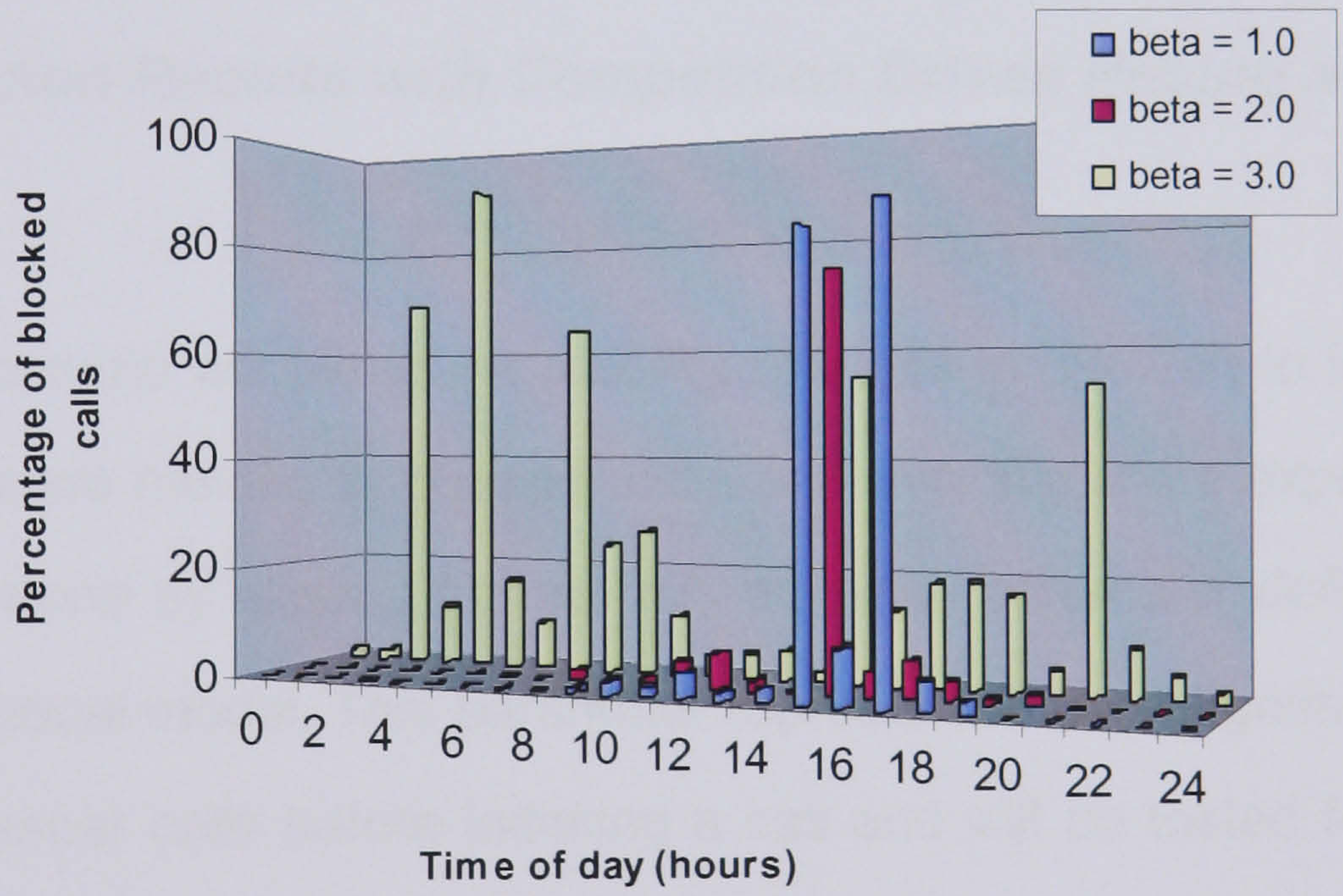

Figure F-4 *Distribution of percentage of blocked calls as a function of time*

Figure F-5 shows the fluctuations of user demand as they occur at 5-minute intervals without smoothing of the data. The graphs show the rapid increase and decrease in demand during off-peak hours as the price fluctuates and explains the very sharp increase in the number of blocked calls as the elasticity of demand β increases.

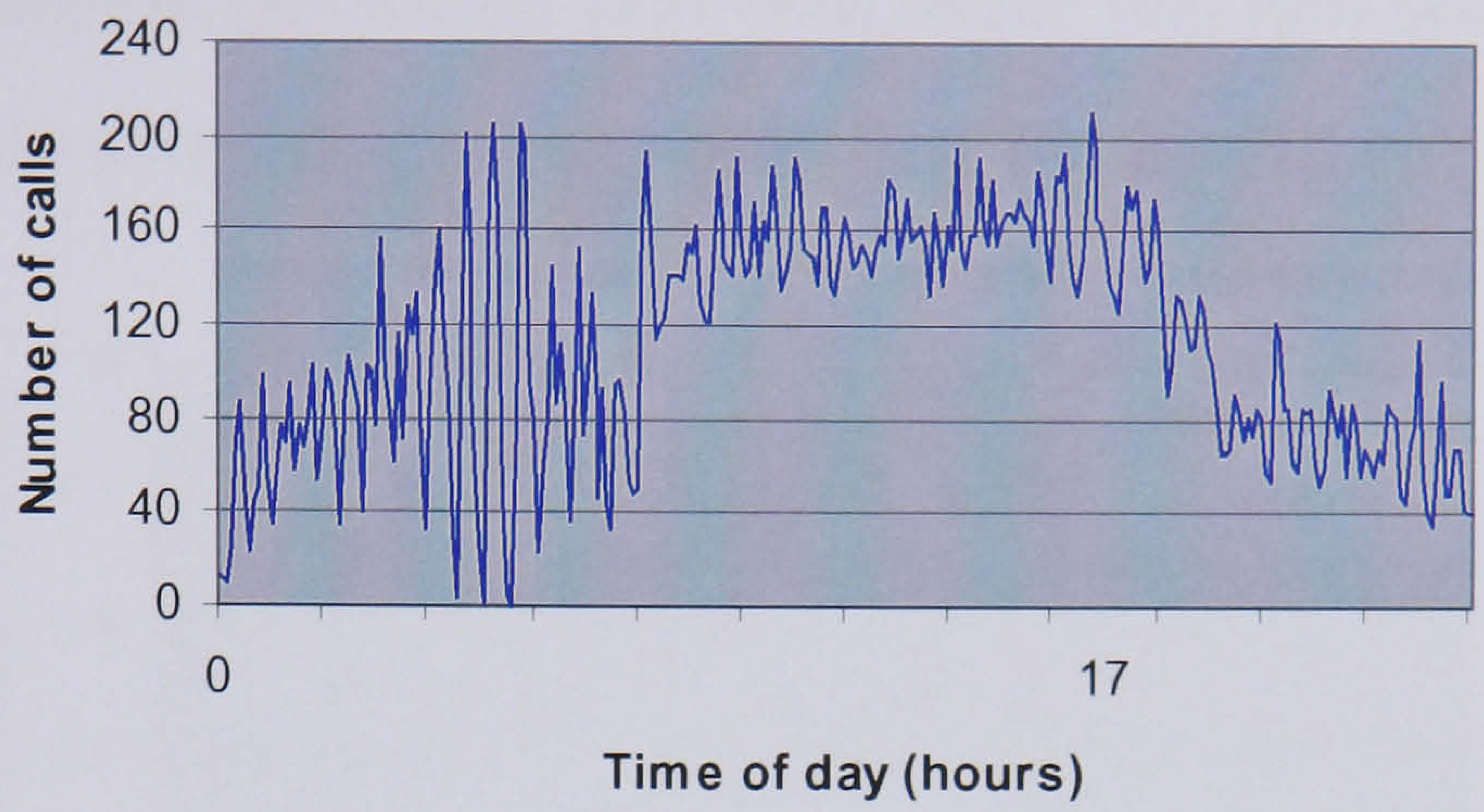

Figure F-5 *Fluctuation of user elastic demand ($\beta = 3.0$) and a non-linear pricing function*

F.3 Simulation Results with Competition Driven Pricing and User Mobility.

This scenario will introduce mobility of users in reaction to the changing price, with users moving to cheaper cells and avoiding more expensive ones. This will be done by altering the mobility elasticity parameter defined as α in the mathematical model. This parameter represents the willingness of users to move to cheaper cells before initiating a call and will be tested for $\alpha \in [0,2]$[56]. As α increases, the willingness of users to move the cheaper cell also increases and, therefore, the relative proportion of users moving to the cheaper cells would grow.

[56] This choice is arbitrary. The value of this parameter can be estimated by network operators by using historic data and conducting further market research (Evans [82]).

F.3.1 Network operator's perspective.

With the introduction of user mobility the simulation predicts a decrease in the total revenue generated with both linear and non-linear pricing function as the elasticity of mobility α increases (see Figure F-6 and Figure F-7).

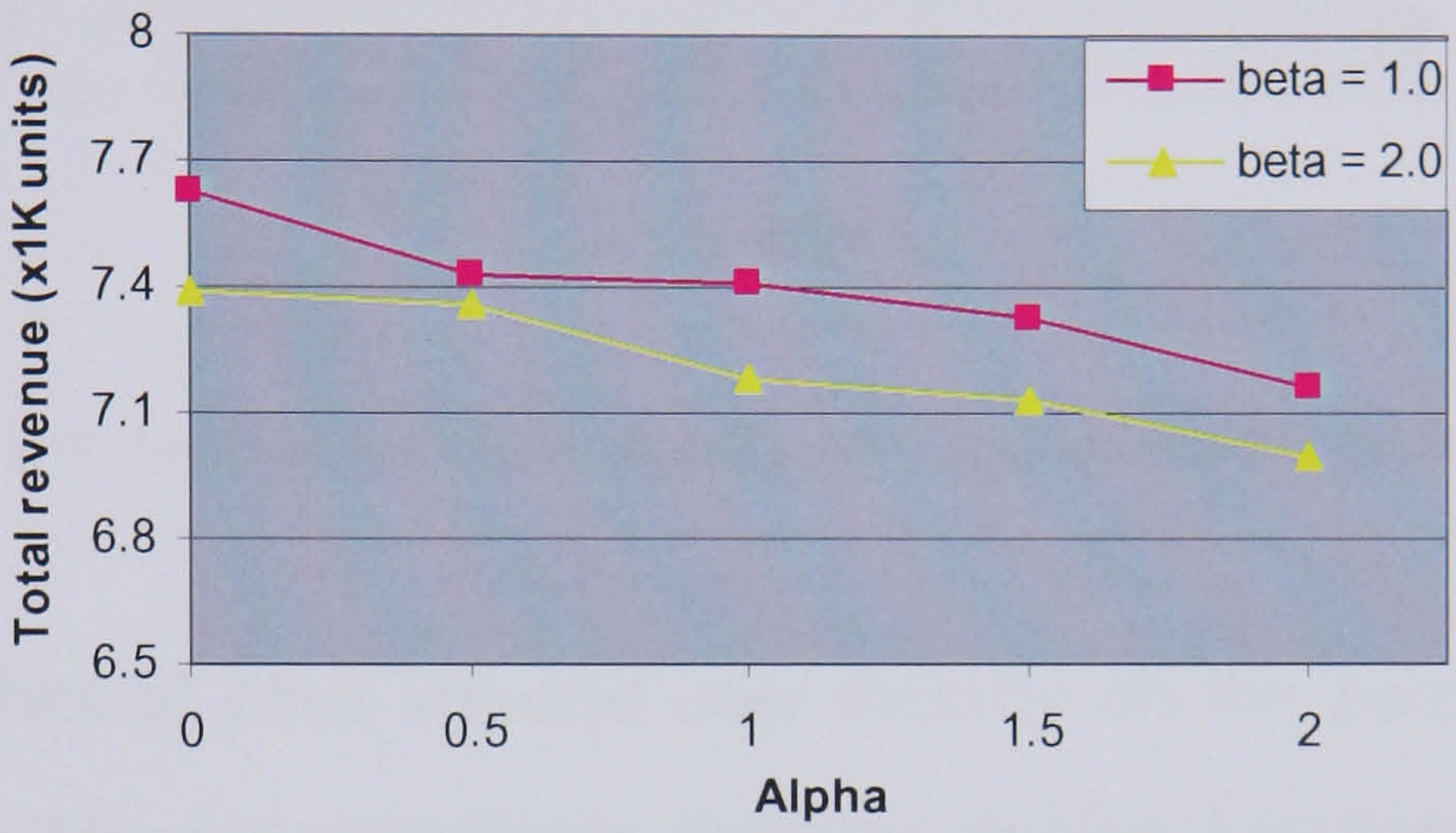

Figure F-6 *Revenue with user mobility and linear pricing function*

This decrease is due to the shift of users to cheaper cells in the network, which results in reduction in the generated revenue. Again the linear and non-linear pricing functions lead to different behaviour in the system. With the linear pricing function the decrease in revenue only becomes significant for $\alpha = 2.0$, at 6% and 5% reduction respectively for $\beta = 1.0$ and $\beta = 2.0$ (see Figure F-6). Network operators would, therefore, only need to take price induced user mobility into account with the linear pricing function, if they estimate the mobility elasticity of users to be greater than 2, *i.e.* $\alpha \geq 2.0$.

With the non-linear pricing function, on the other hand, the reduction becomes significant at 5.5% for $\alpha = 1.0$ and inelastic demand ($\beta = 1.0$) and at 5.7% reduction for $\alpha = 0.5$ and unit elastic demand ($\beta = 2.0$) (see Figure F-7). Therefore, with the non-linear pricing function the network operator has to take into account user mobility as soon as dynamic pricing is introduced.

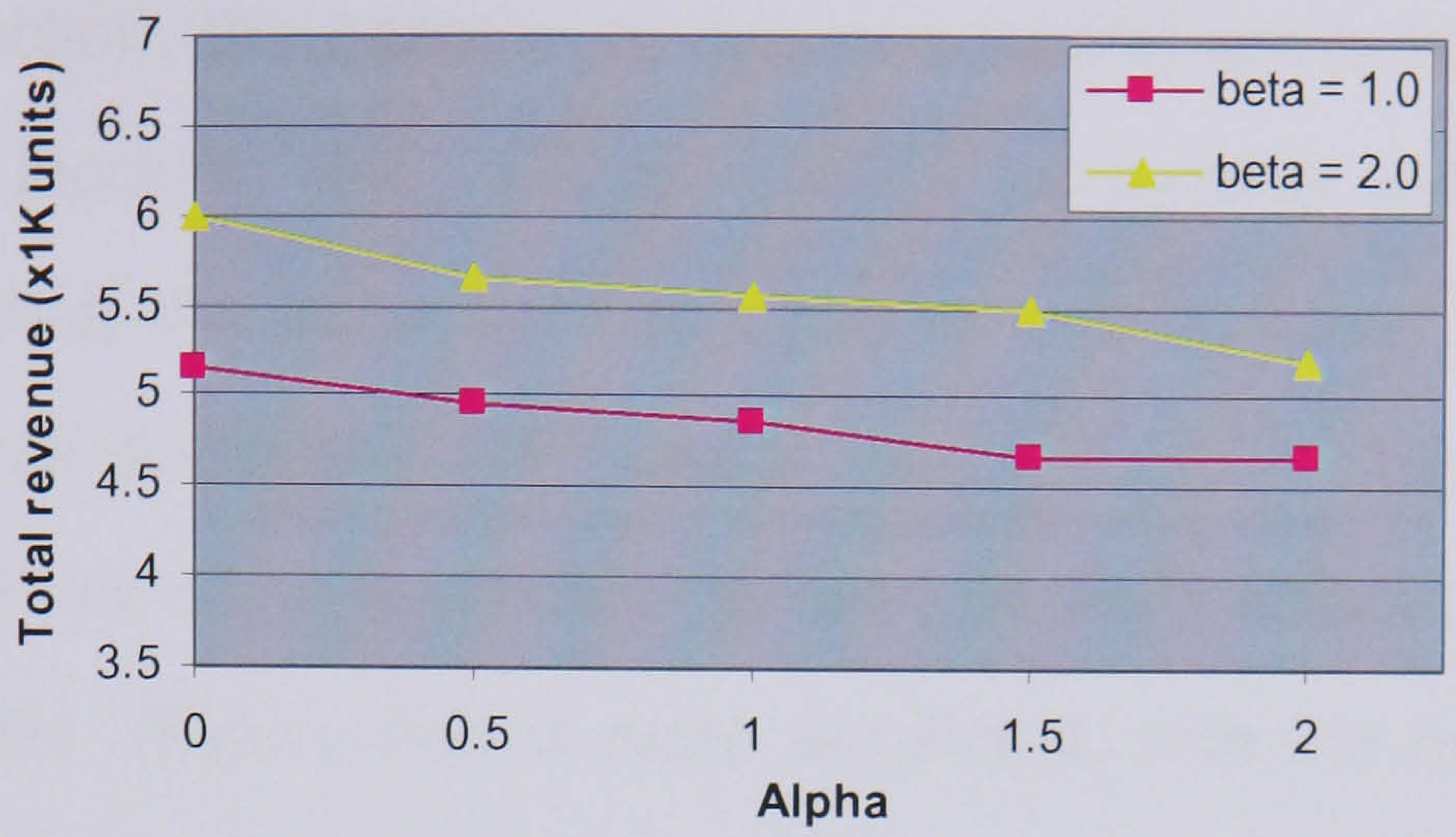

Figure F-7 *Revenue with user mobility and non-linear price function*

The effect of price induced user mobility on the percentage of blocked calls with linear and non-linear dynamic pricing functions can be seen in Figure F-8 and Figure F-9 respectively.

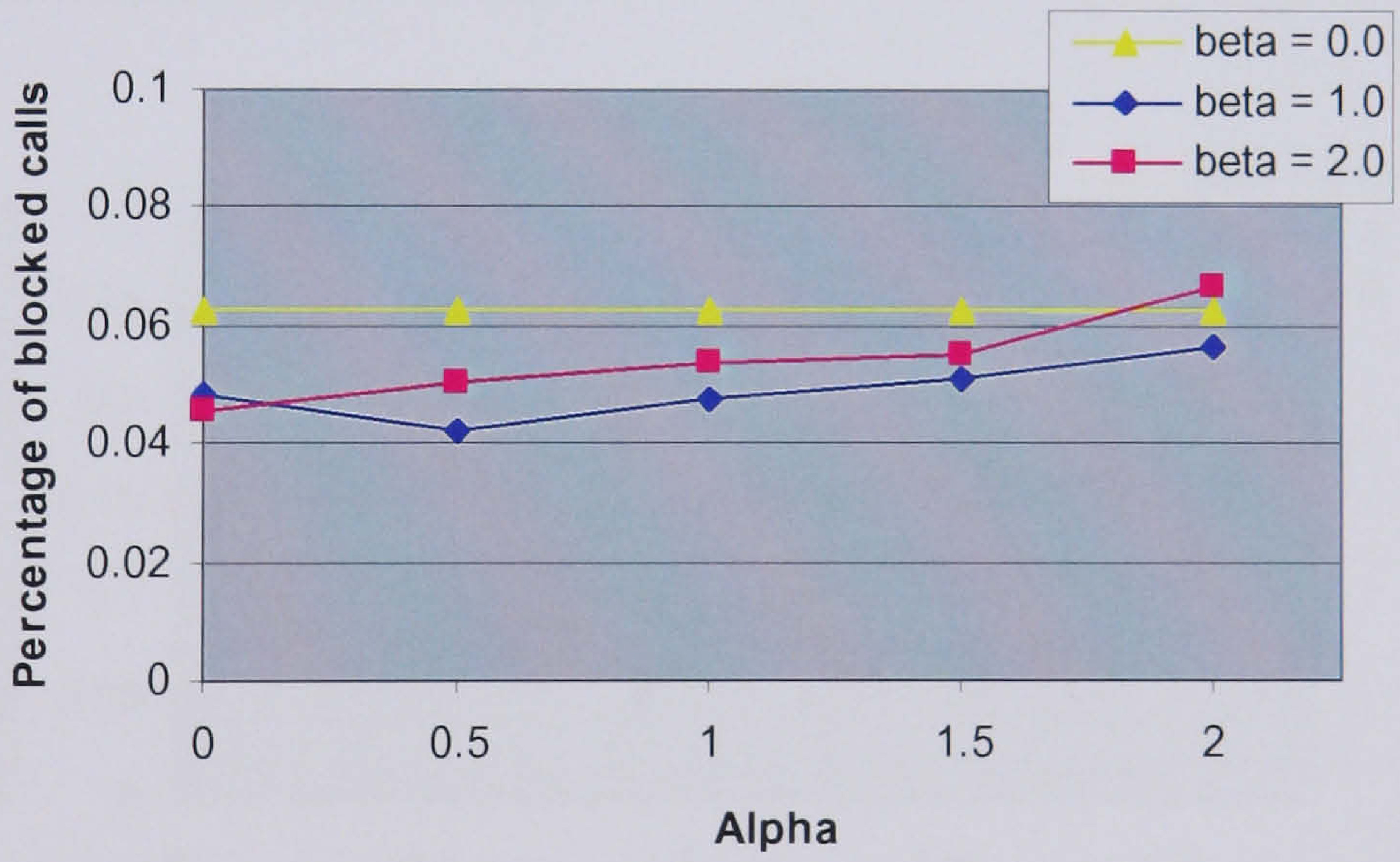

Figure F-8 *Percentage of blocked calls with mobility and linear pricing function*

With the linear pricing function (Figure F-8) call blocking decreases by 13% when mobility is first introduced and demand is inelastic ($\alpha \leq 1.0$ and $\beta = 1.0$) but then increases steadily until there is 45% more blocking with unit elastic demand ($\beta = 2.0$). This increase is due to the increase in the proportion of people moving to cheaper cells and saturating the available capacity, before the price increases and drives them away. However, in the case of the linear pricing function, the benefits of dynamic pricing are only outweighed by the increased blocking due to user mobility for mobility elasticity $\alpha > 1.5$. This suggests that, the linear *ad hoc* dynamic pricing could offer a reduction in expected call blocking.

The effect of the non-linear dynamic pricing function on the percentage of blocked calls (Figure F-9) is more significant, with the overall percentage of blocked calls increasing for both inelastic and unit elastic demand, due to cheaper cell saturation. The increase is very significant for inelastic demand ($\beta = 1.0$) reaching up to 49% for $\alpha = 2.0$. However, with unit elastic demand ($\beta = 2.0$) the increase is not significant for $\alpha \leq 1.0$ at 0.1% increase. As the mobility elasticity α increases this difference becomes more significant reaching 35% increase for $\alpha = 2.0$.

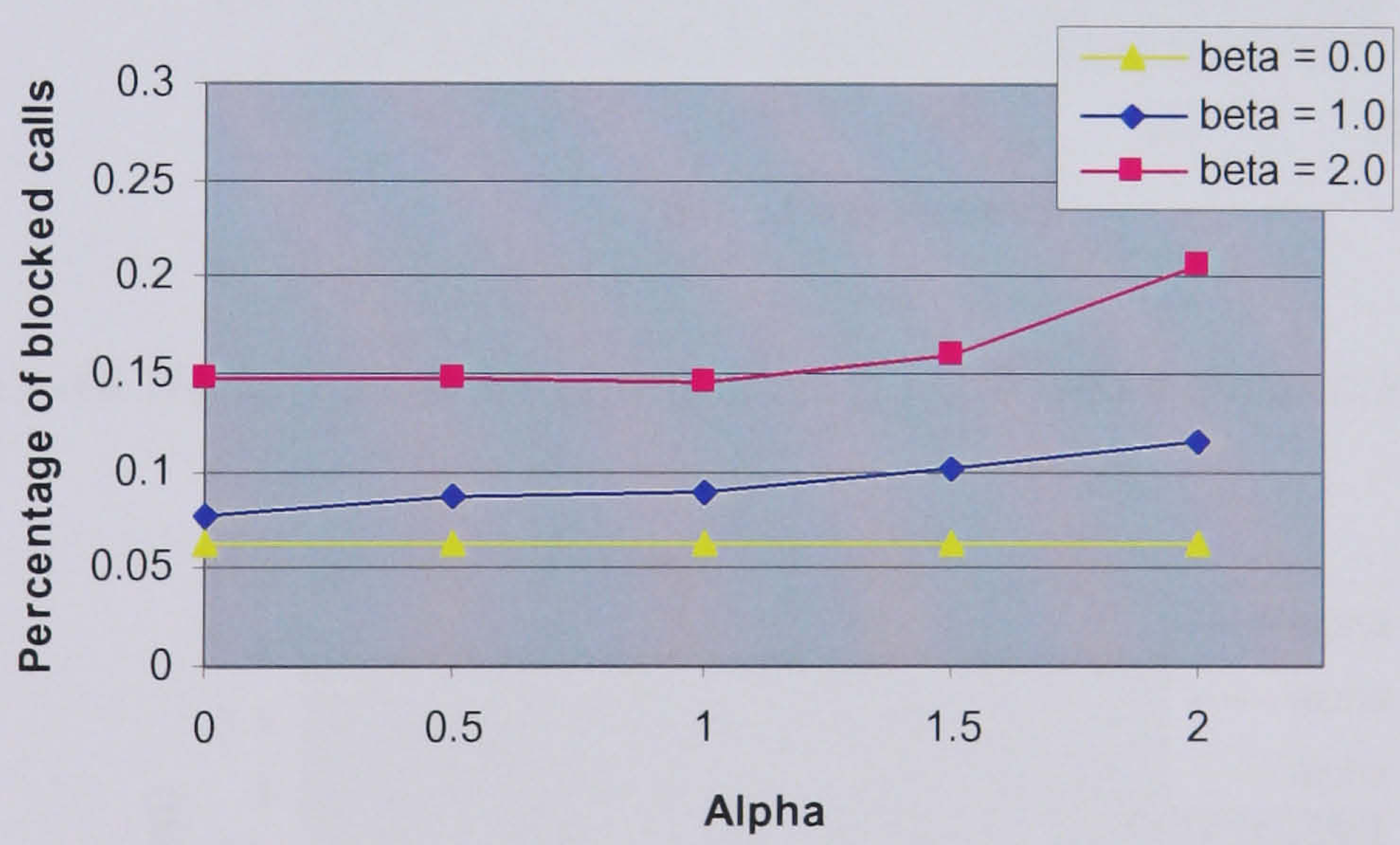

Figure F-9 *Percentage of blocked calls with mobility and non-linear pricing function*

Therefore, although call blocking in the network can increases as the proportion of users moving to cheaper cells increases, this effect is only significant for mobility elasticity $\alpha > 1.5$ with the linear pricing function. With the non-linear pricing function the effect becomes significant for mobility elasticity $\alpha > 0.5$ and so, network operators would have to be more cautious when applying this form of dynamic pricing.

The effect of price induced user mobility elasticity α on the variance of the load in the network is shown in Figure F-10 and Figure F-11 for inelastic demand ($\beta = 1.0$) with the linear and non-linear pricing function respectively. The introduction of mobility ($\alpha > 0.0$) leads to a decrease in the intensity of peak calls for both pricing functions, with the decrease with non-linear pricing function being more significant (Figure F-11).

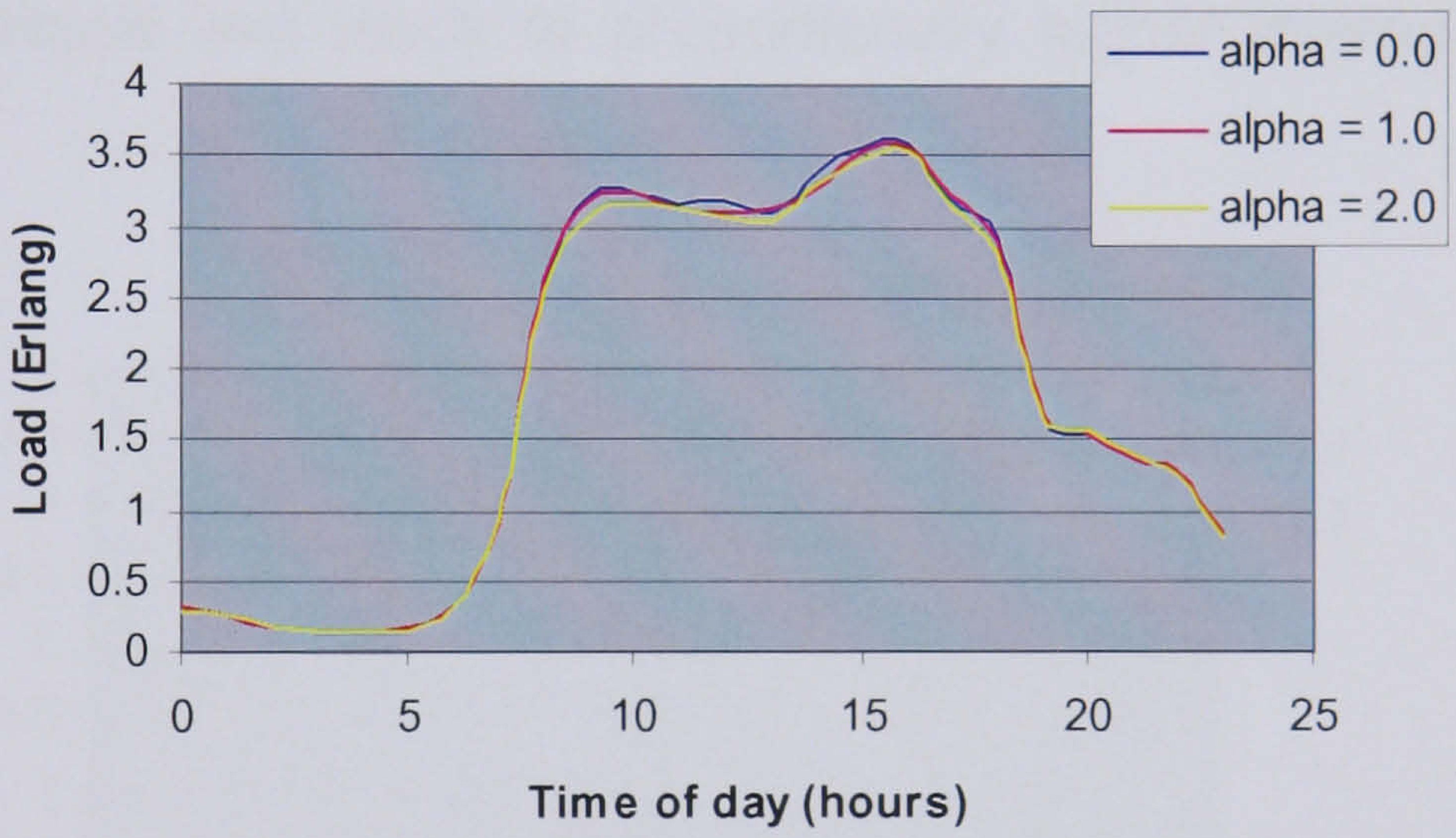

Figure F-10 *Traffic load with user mobility, $\beta = 1.0$ and a linear pricing function*

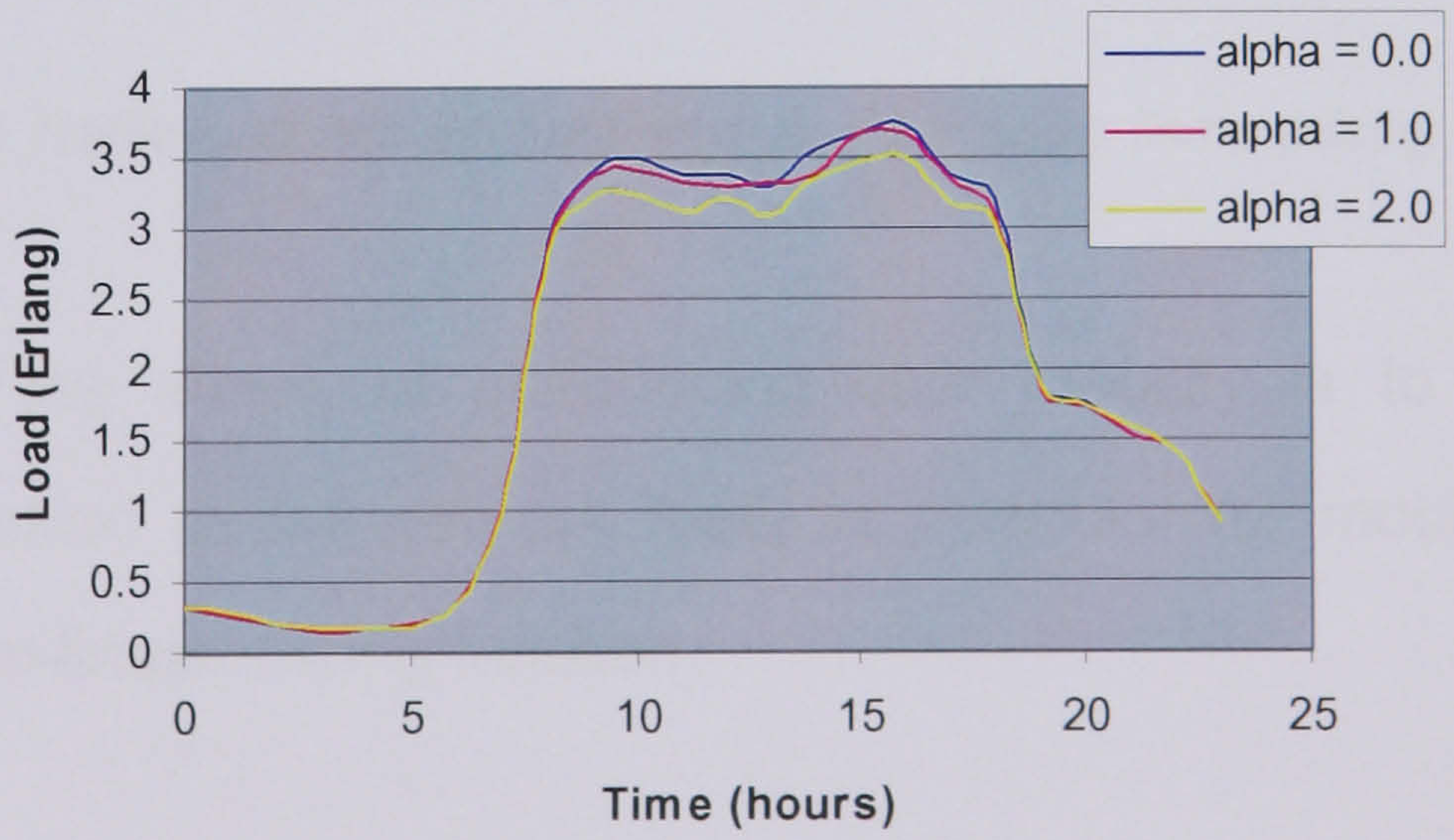

Figure F-11 *Traffic load with user mobility, $\beta = 1.0$ and a non-linear pricing function*

The effect of price induced user mobility elasticity α on the variance of the load in the network for unit elastic demand ($\beta = 2.0$), is plotted in Figure F-12 and Figure F-13.

The introduction of user mobility again leads to a more even distribution of the network load as a function of time. As α increases the effect on the off-peak network load is insignificant. There is, however, a sizeable reduction (up to 0.5 Erlang for the non-linear function, Figure F-13) in the expected load in the network during peak hours. This is reduction due to the overall increase in the average price in the network, as the proportion of users shifting to less busy cells increases and leads to proportionally higher average price (see Figure F-17).

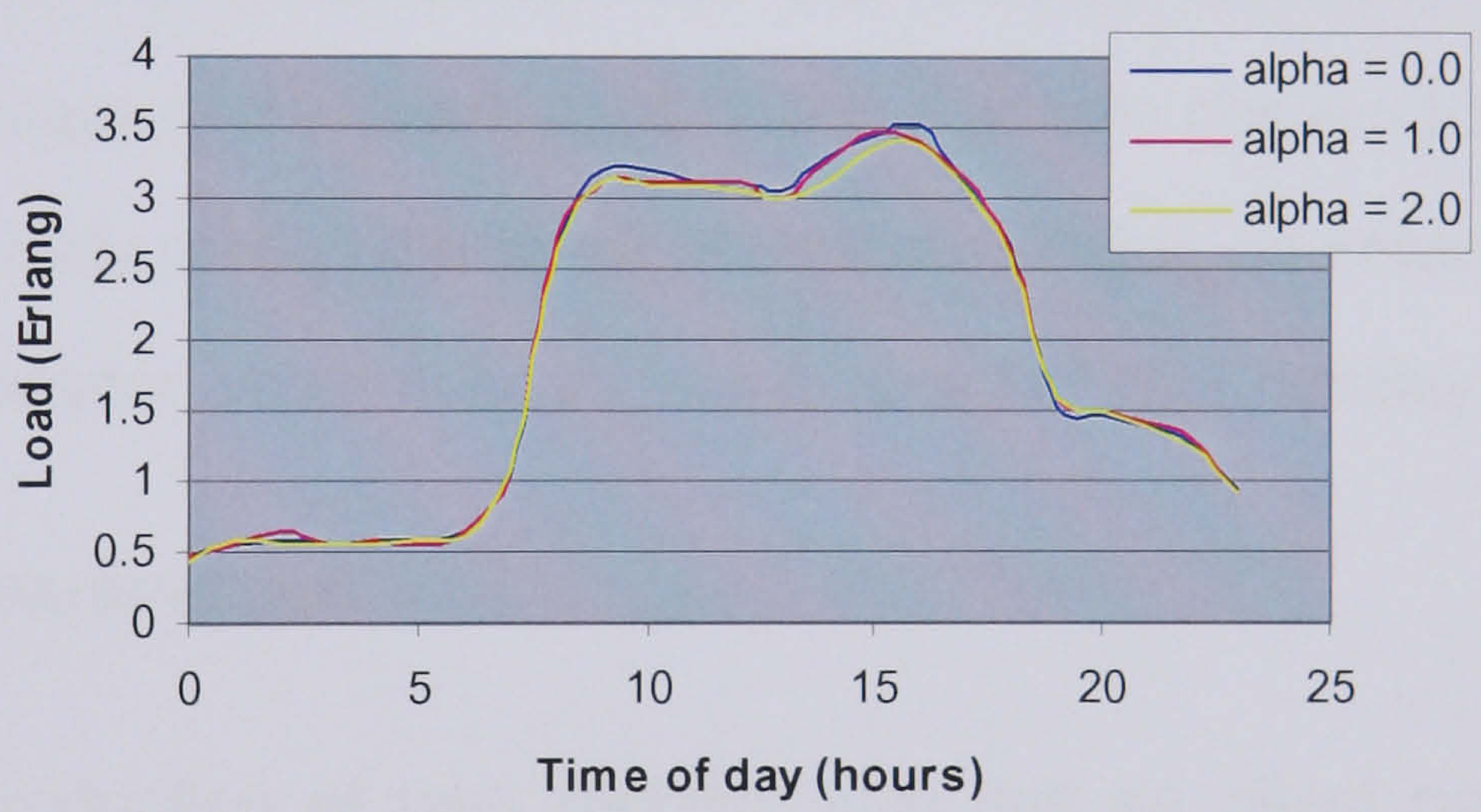

Figure F-12 *Traffic load with user mobility,* $\beta = 2.0$ *and a linear pricing function*

Therefore the effect, of intrioducing user mobility is to smooth the temporal distribution of the network load, in particular for mobility elasticity $\alpha \geq 1.0$ and a non-linear pricing function.

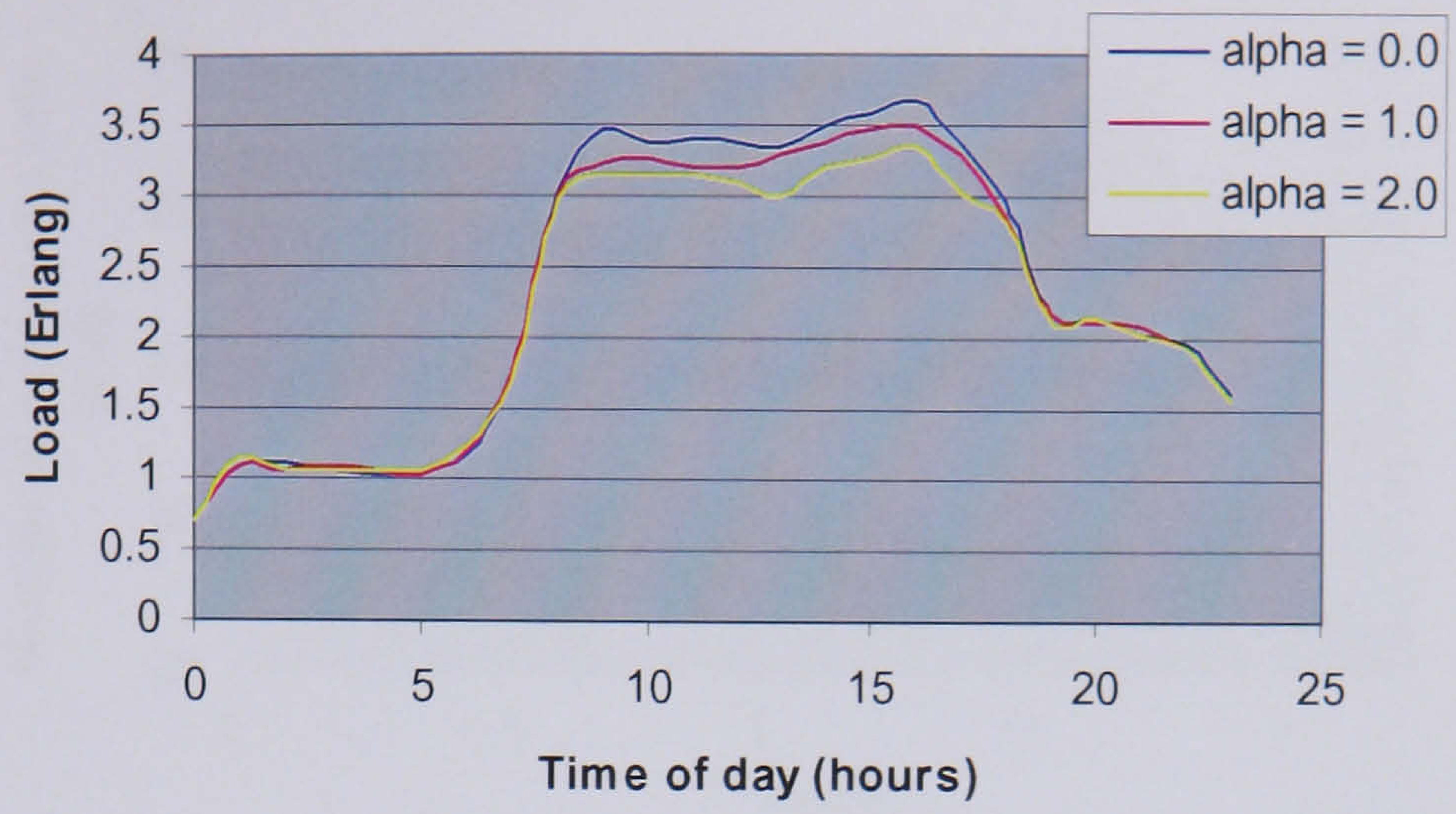

Figure F-13 *Traffic load with user mobility,* $\beta = 1.0$ *and a non-linear pricing function*

These results show that the introduction of user mobility as a function of the price in the network could lead to a reduction in the amount of revenue generated for the network operator (for mobility elasticity $\alpha > 0.5$) and a potential increase to the percentage of blocked calls (for $\alpha > 1.5$). Therefore, in order to precisely predict the effect of dynamic pricing on network performance a network operator would need to take account of user mobility.

F.3.2 User perspective.

The introduction of user mobility also has an effect on the welfare of users by affecting the number of successful calls and the weighted average price in the network.

In Figure F-14 and Figure F-15 the number of serviced calls for different values of α are compared.

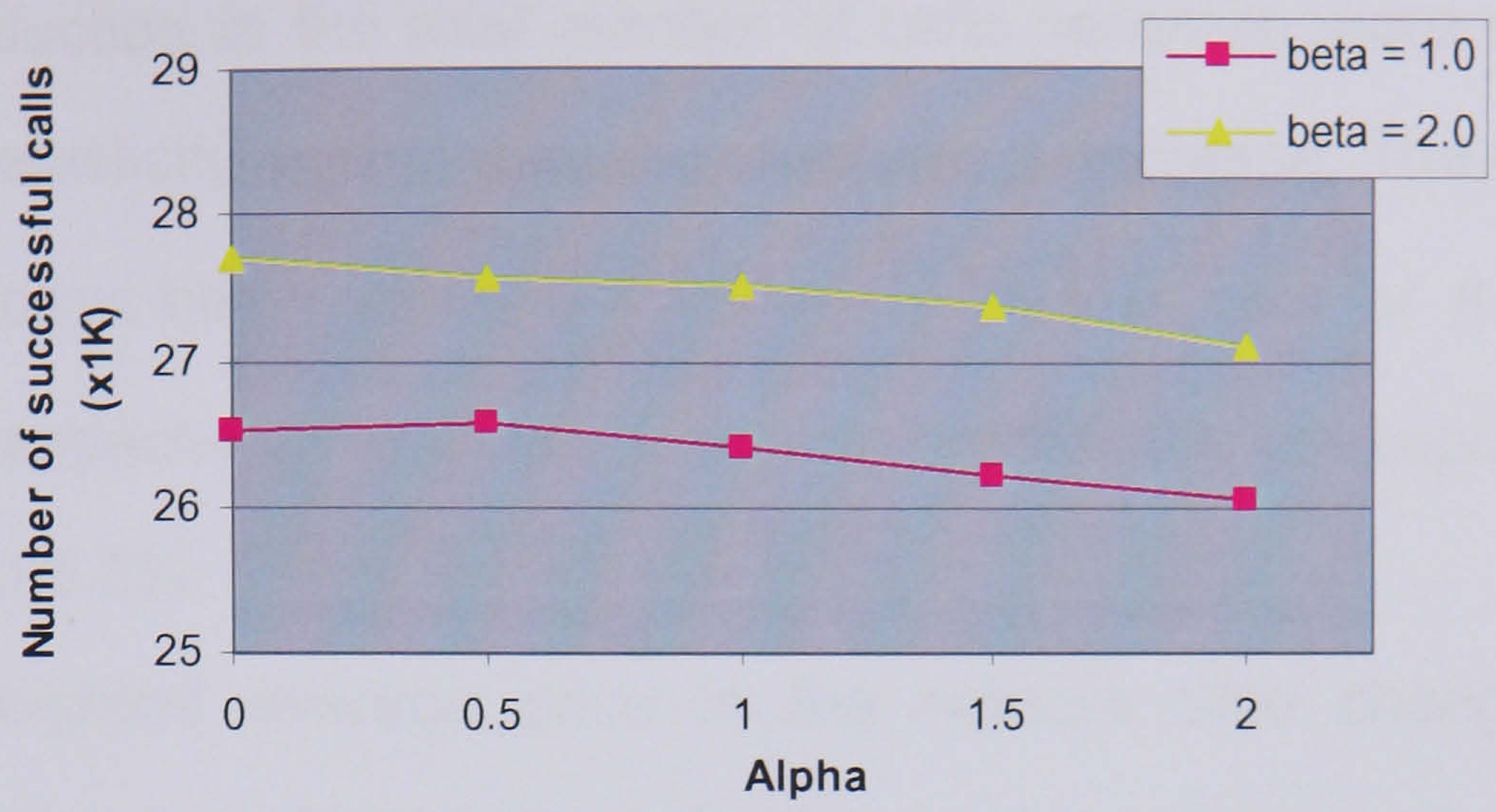

Figure F-14 *Total number of successful calls with mobility and a linear price function*

Generally, the introduction of mobility leads to fewer calls being made with both linear and non-linear pricing functions. This is due to the reduction of calls in peak hours, as α increases, which is not compensated by an increase in calls during off peak hours. The only exception is the increase in the number of successful calls by 0.1% with a linear pricing function, $\alpha = 0.5$ and inelastic demand ($\beta = 1.0$).

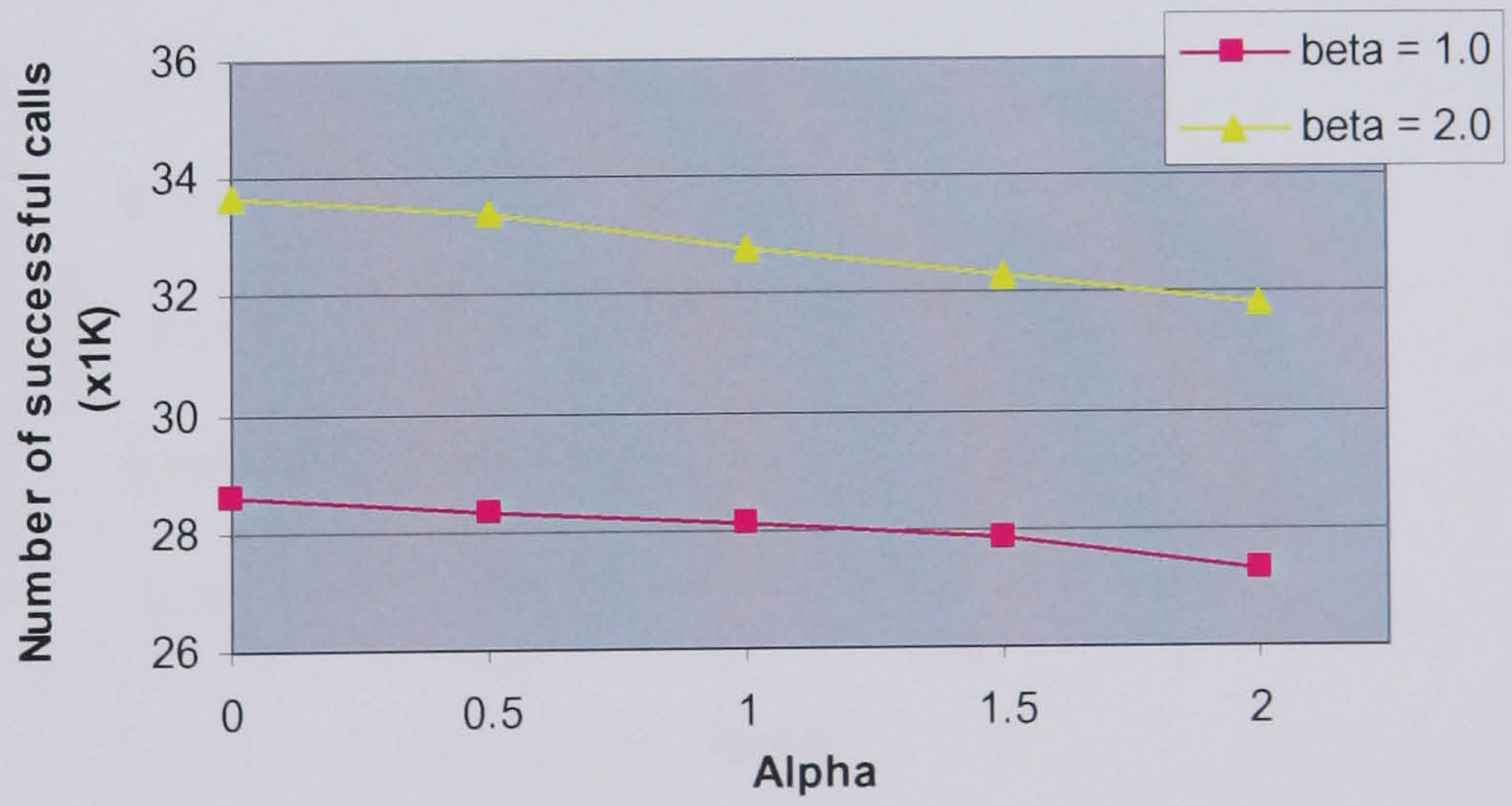

Figure F-15 *Total number of successful calls with mobility and a non-linear price function*

The reduction in the total number of calls becomes more significant as the mobility elasticity α and demand elasticity β increase. The linear pricing function performs better with only 1.8% and 2.1% less calls for $\beta = 1.0$, $\beta = 2.0$ and $\alpha = 2.0$ respectively. With the non-linear function the corresponding values are 4.4% and 5.5%.

The weighted average price in the network also changes with the introduction of user mobility for both the linear and non-linear price functions.

With the linear pricing function the average price in the network decreases by 1.3% as user mobility is introduced for inelastic demand ($\beta = 1.0$) (see Figure F-16). The decrease diminished as α increases leading to - 0.1% percent decrease in the price for $\alpha = 2.0$, which is insignificant, compared to the situation without user mobility.

For unit elastic demand ($\beta = 2.0$), the price remains constant as the mobility elasticity α increases with a maximum 0.6% increase at mobility elasticity $\alpha = 2.0$ (which is insignificant).

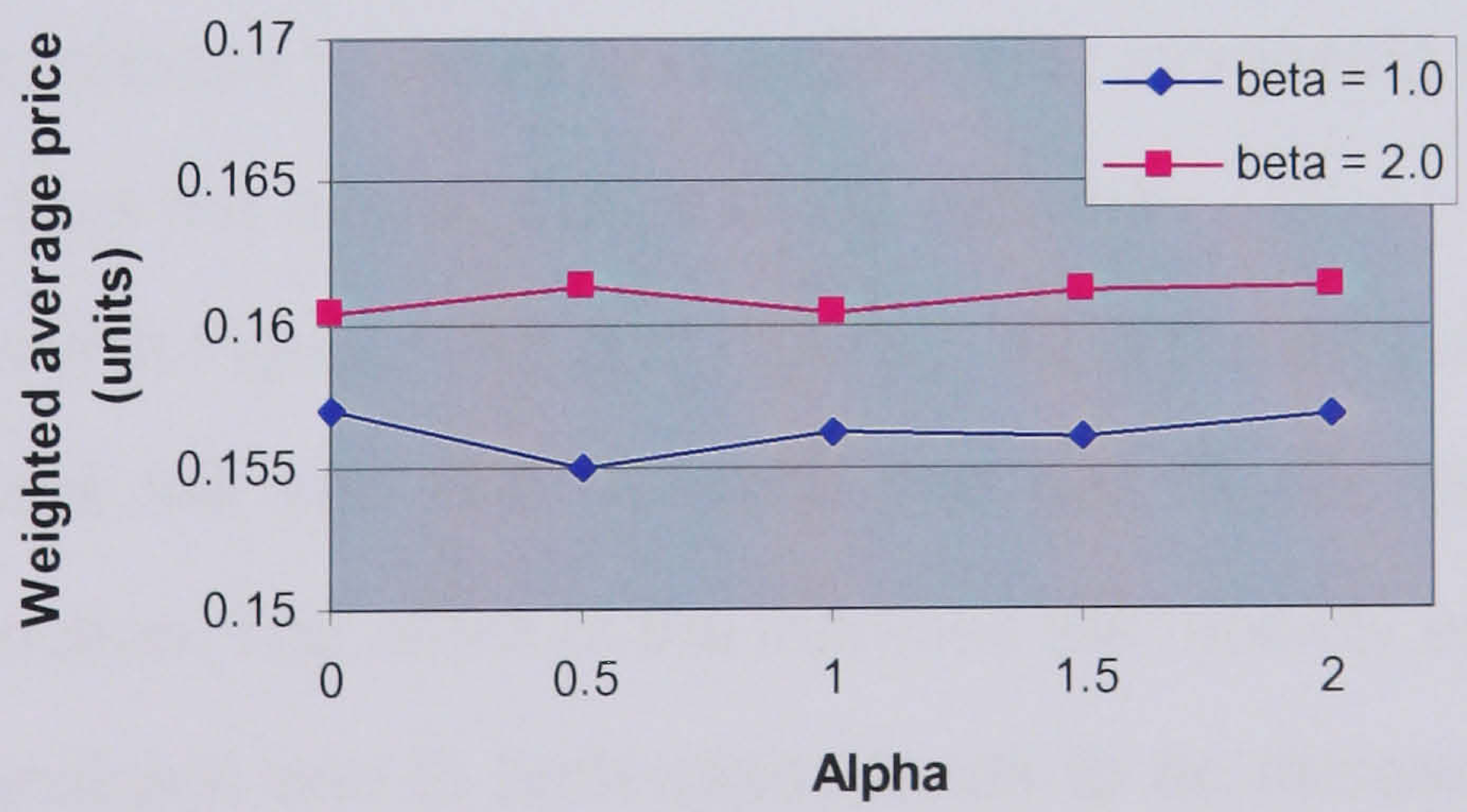

Figure F-16 *Weighted average price with mobility and a linear pricing function*

The situation is more dynamic with the non-linear pricing function. In this case, the price steadily increases as α increases with the inelastic demand

($\beta = 1.0$) reaching up to 8.9% more, for mobility elasticity $\alpha = 2.0$. For unit elastic demand ($\beta = 2.0$), however, the introduction of user mobility leads to an initial drop in the average price in the network by 3% (mobility elasticity $\alpha \leq 1.0$).

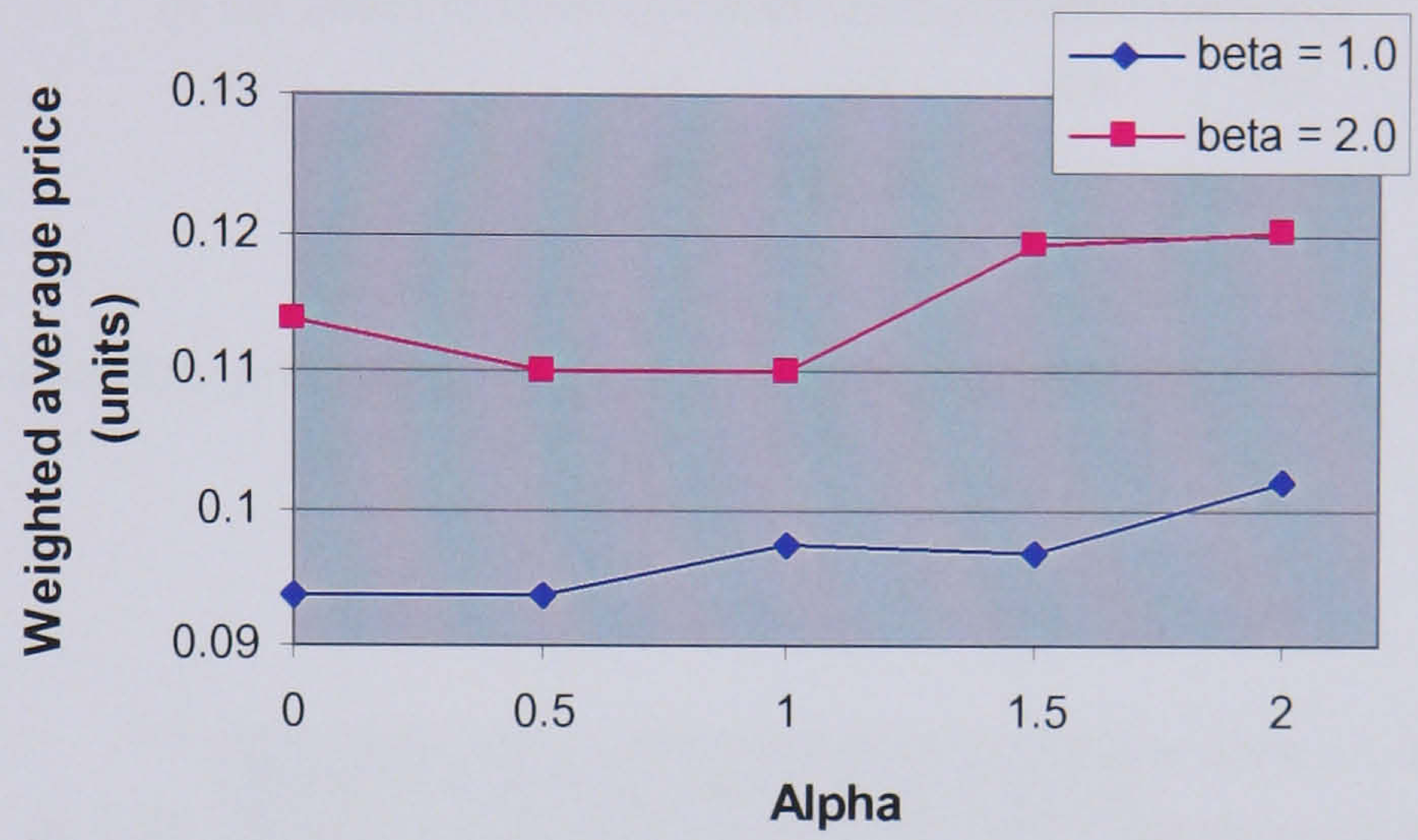

Figure F-17 *Weighted average price with mobility and a non-linear pricing function*

As the mobility elasticity increases to $\alpha = 2.0$, the average price in the network also increases by up to 5.7% above the average price without user mobility. This price increase is due to the relative increase in the proportion of users who choose to move to cheaper cells, increasing the overall price in the cells and thus the average price in the network.

Shown in Figure F-18 and Figure F-19 is the probability distribution of the prices users will see with inelastic and unit elastic demand and the linear pricing function. The effect of the increase the mobility elasticity of users (α) is very insignificant and in both cases leads to an increase in the probability of users seeing the minimum and the maximum price in the network.

Figure F-18 *Price probability distribution with user mobility* $\beta = 1.0$ *and a linear pricing function*

Figure F-19 *Price probability distribution with user mobility* $\beta = 2.0$ *and a linear pricing function*

Plotted in Figure F-20 and Figure F-21 is the effect of the user mobility elasticity α on the probability distribution of expected prices with the non-linear pricing function. As was the case with the linear price function, the effect of the increase the mobility elasticity of users (α) is very insignificant and in both cases leads to an increase in the probability of users seeing the minimum and the maximum price in the network, rather than the intermediate prices.

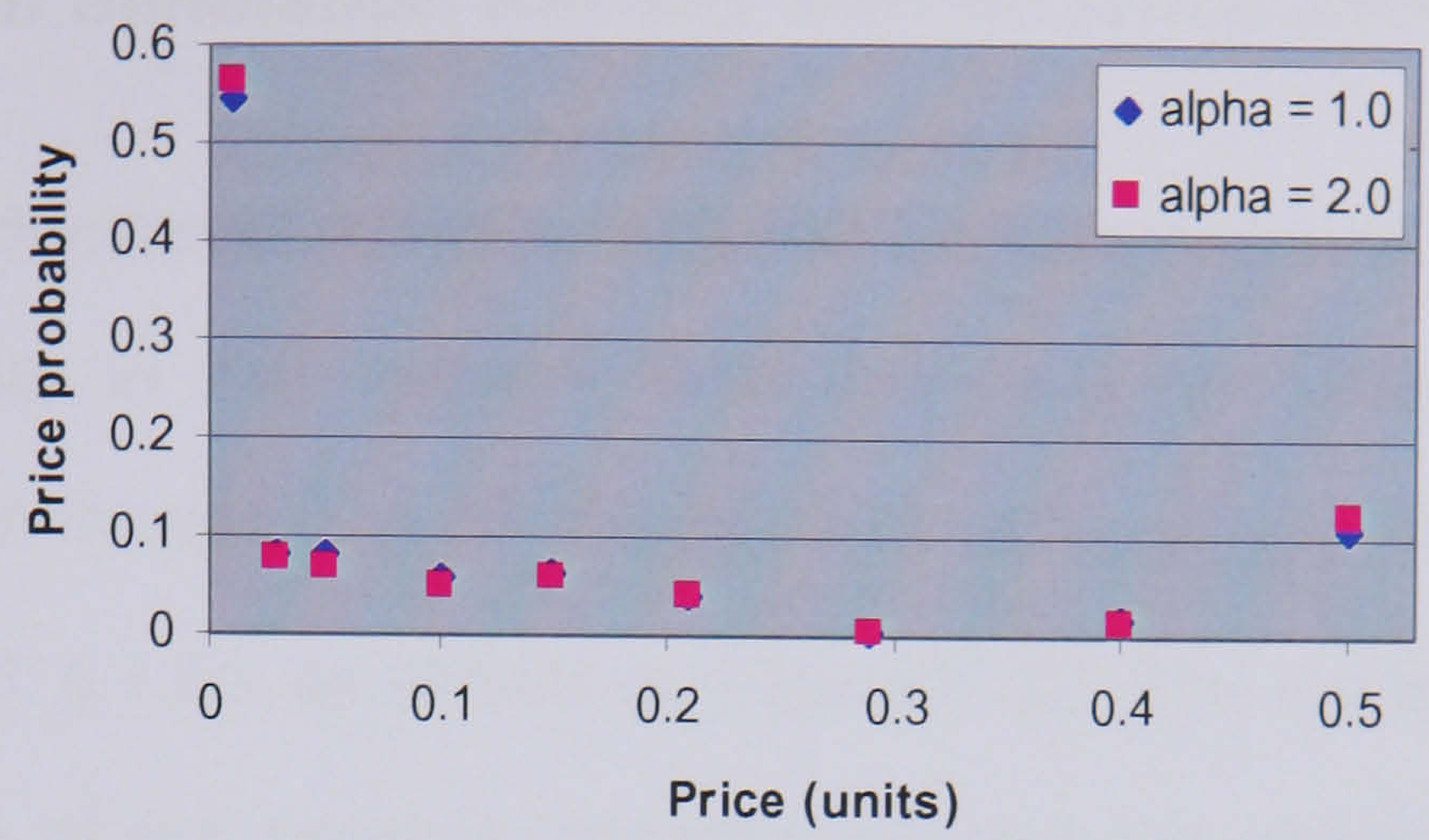

Figure F-20 *Price probability distribution with user mobility* $\beta = 1.0$ *and a non-linear pricing function*

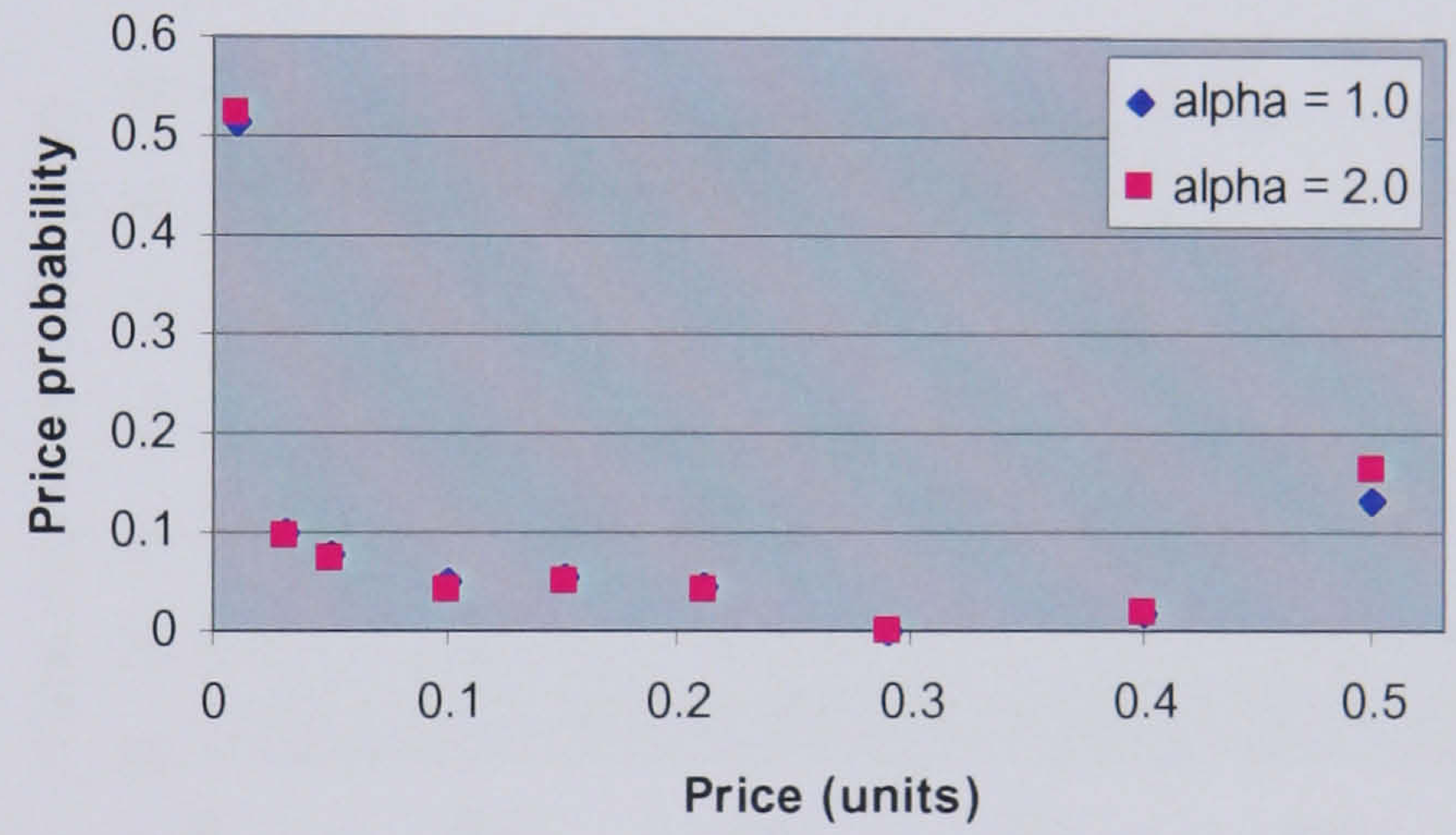

Figure F-21 *Price probability distribution with user mobility* $\beta = 2.0$ *and a non-linear pricing function*

These results suggest that taking into account user mobility due to different prices in the network, combined with dynamic pricing, could potentially lead to a reduction in the number of successful calls in the network. It can also lead to an increase in the expected average price in the network for $\alpha > 1.0$ if non-linear pricing function is used.

F.4 Additional Simulation Results with Revenue Attainment Pricing.

The introduction of revenue attainment pricing leads to an increase in the successful calls in the network with low revenue attainment preference ($\varepsilon = 0.1$) with an increase by 8% for inelastic demand $\beta = 1.0$ and 35% for unit elastic demand $\beta = 2.0$ as shown by Figure F-22. The increase diminishes as the preference of the network operator for revenue attainment ε increases (increase by 5% for inelastic demand $\beta = 1.0$ and 15% for unit elastic demand $\beta = 2.0$. Overall, the total number of successful calls is higher than the number of successful calls in the system without dynamic pricing.

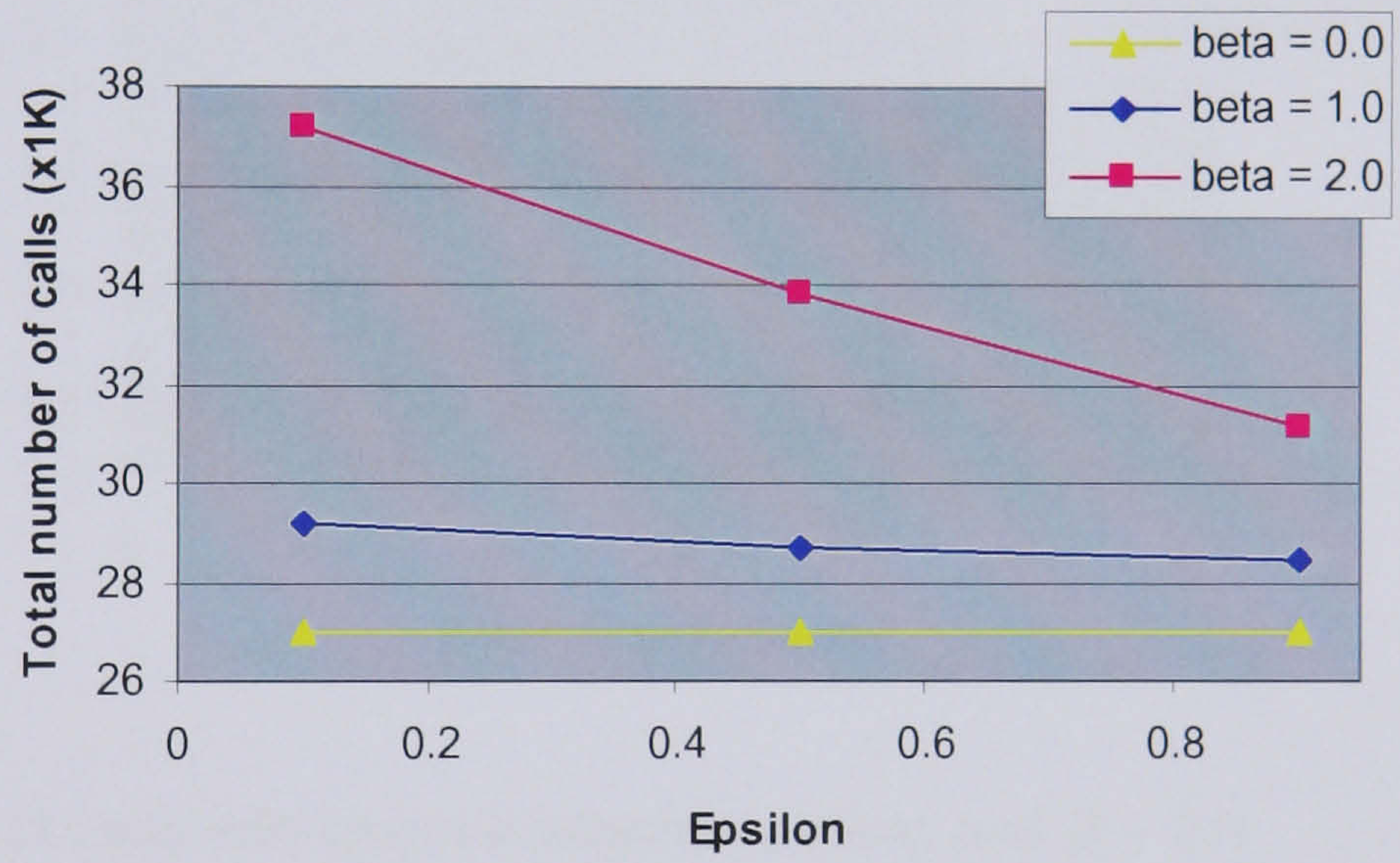

Figure F-22 *Total number of successful calls with revenue attainment prices*

The effect of the preference for revenue attainment ε on the load in the network is plotted in Figure F-23 and Figure F-24. The graphs show that as the preference for revenue attainment ε increases (and, therefore, the preference for maximum capacity utilisation ϕ decreases) the overall load in the network also decreases. Significantly, however, the load in the network with dynamic pricing is higher than the load without dynamic pricing, regardless of demand elasticity.

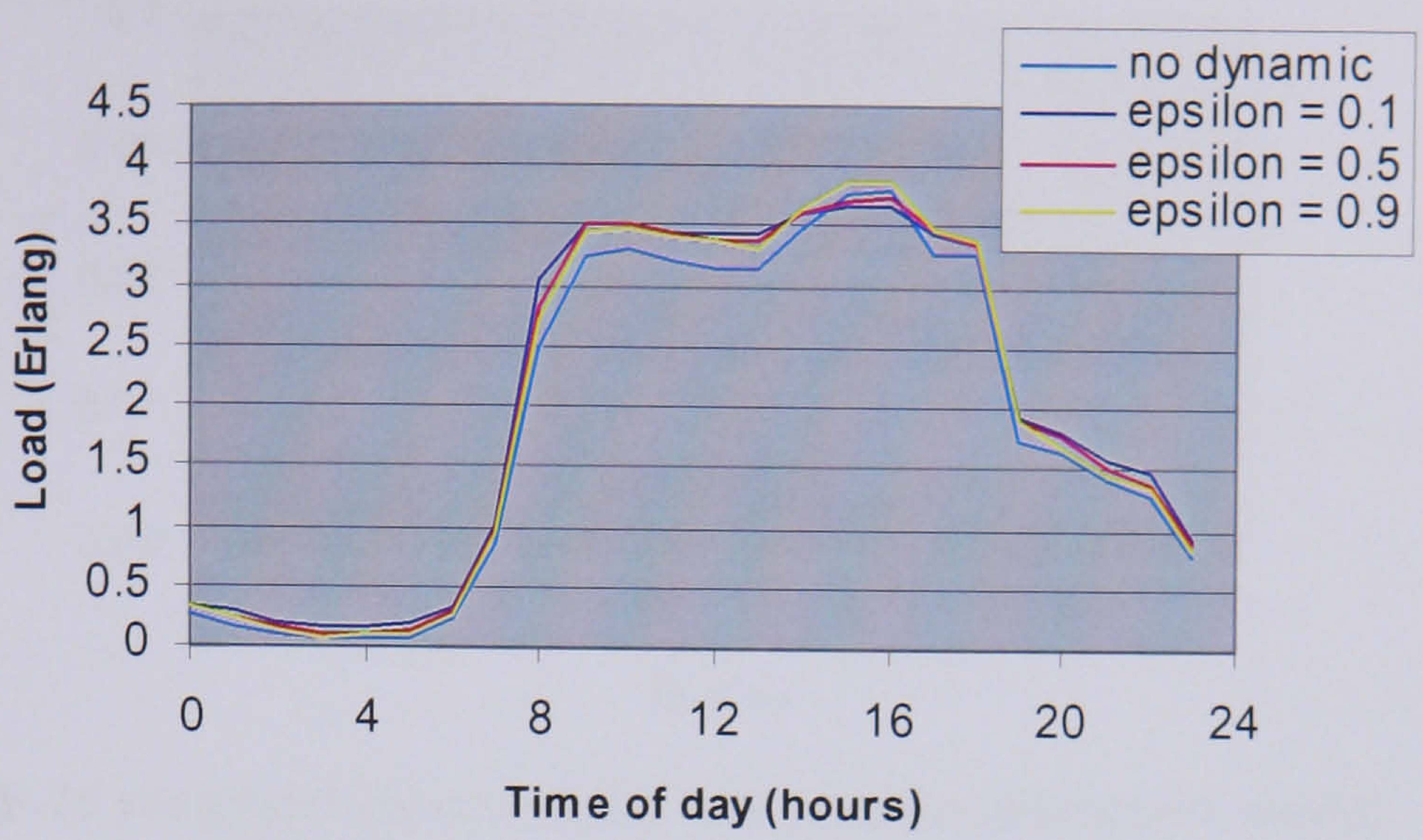

Figure F-23 *Load with revenue attainment prices and* $\beta = 1.0$

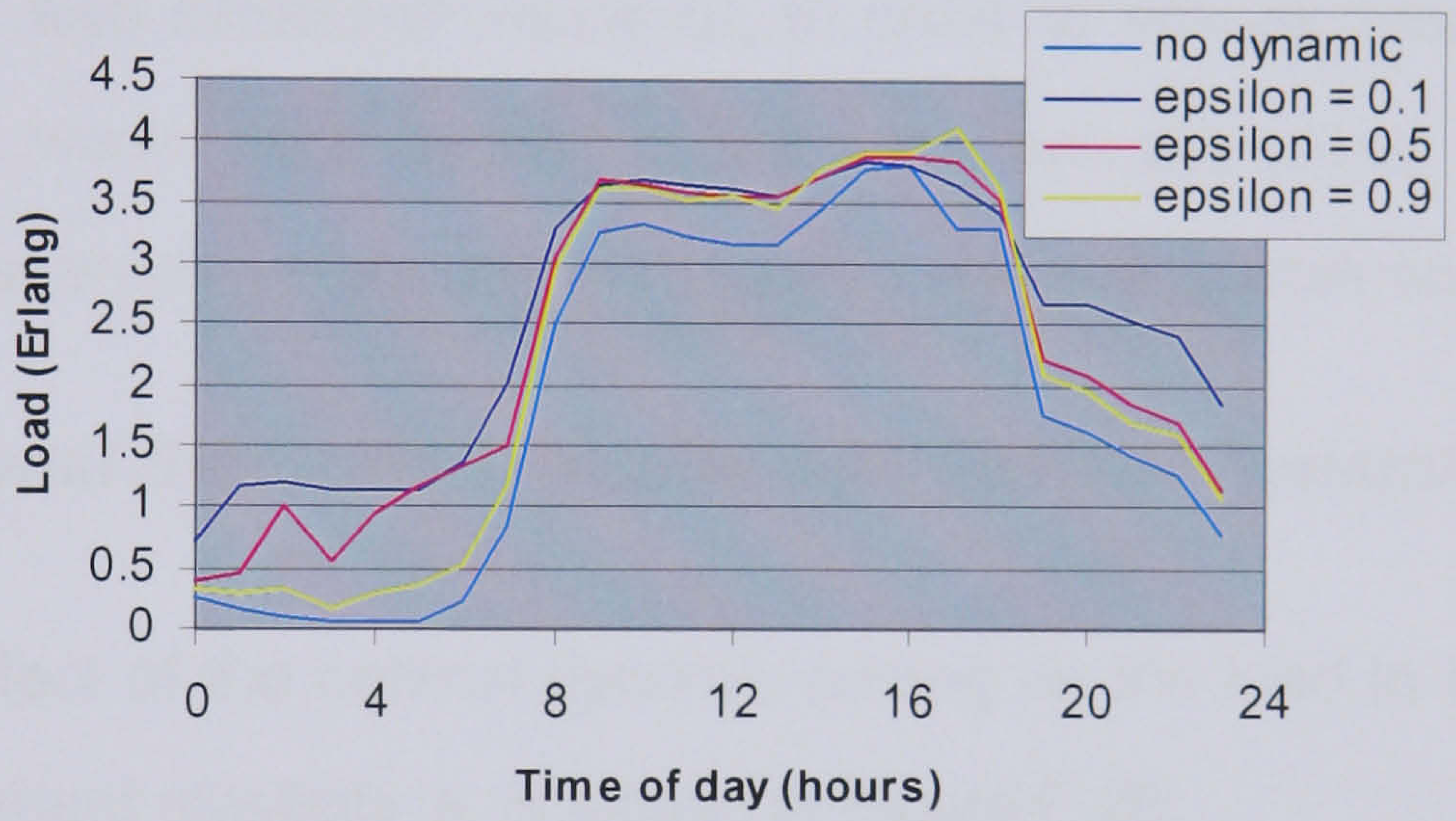

Figure F-24 *Load with revenue attainment prices and* $\beta = 2.0$

This shows that dynamic pricing leads to better utilisation of the available capacity by reducing pieces when network capacity is under-utilised and stimulating demand

The weighted average prices in the network are shown in Figure F-25. Overall the average price in the network is significantly lower (up to 35%) than without dynamic pricing (0.13 units)[57].

[57] Calculated using the Cellnet table (Occasional Caller +) in Appendix B.

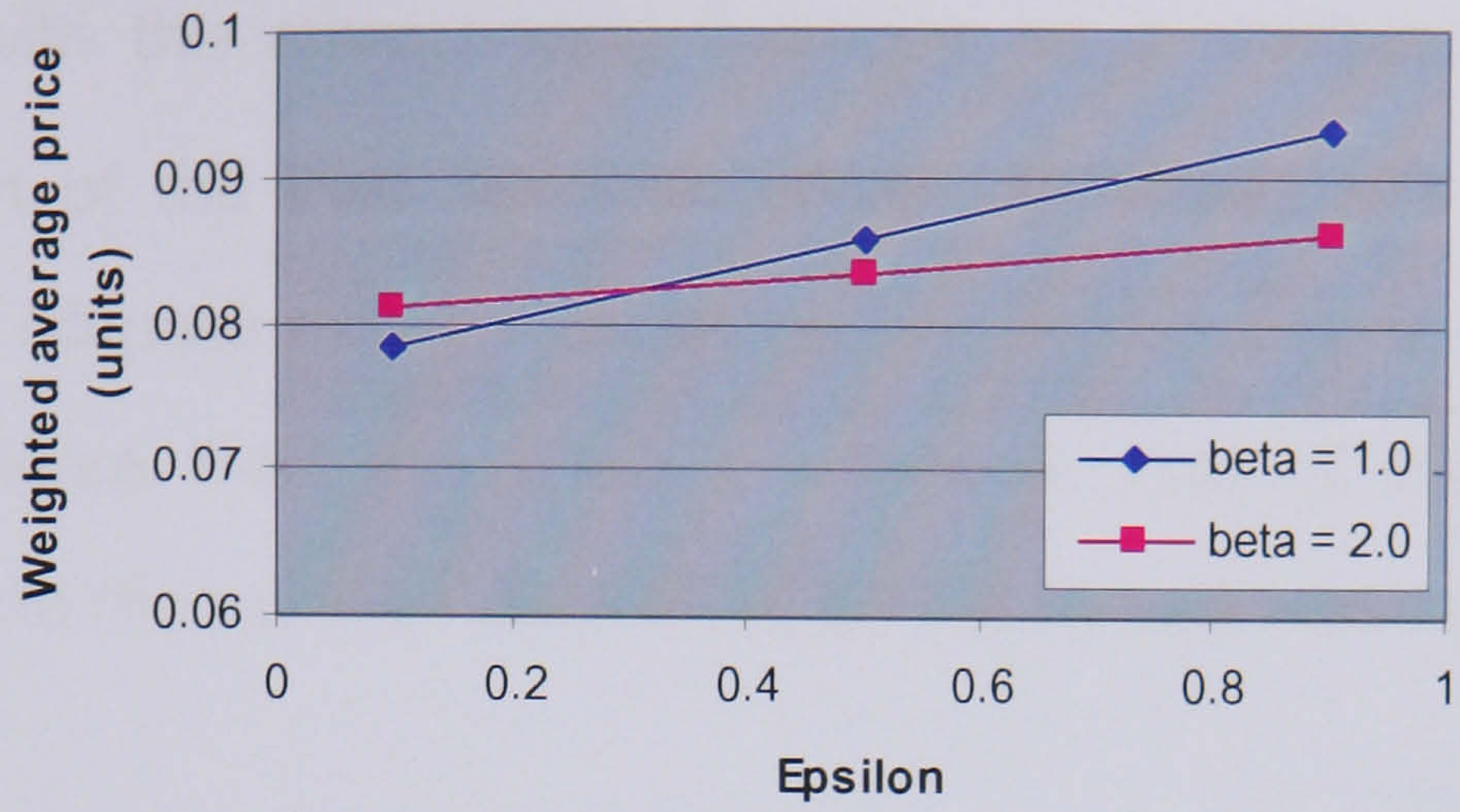

Figure F-25 *Weighted average price with revenue attainment model*

As the preference for revenue attainment increases, the average price in the network also increases. However, in order to ensure maximum capacity utilisation, it would be necessary to keep the average prices lower than the prices necessary to ensure that the required revenue is attained.

F.5 Additional Simulation Results with Optimal Dynamic Pricing.

The effect of the optimal dynamic pricing on the load in the network for different demand elasticity β is plotted in Figure F-26.

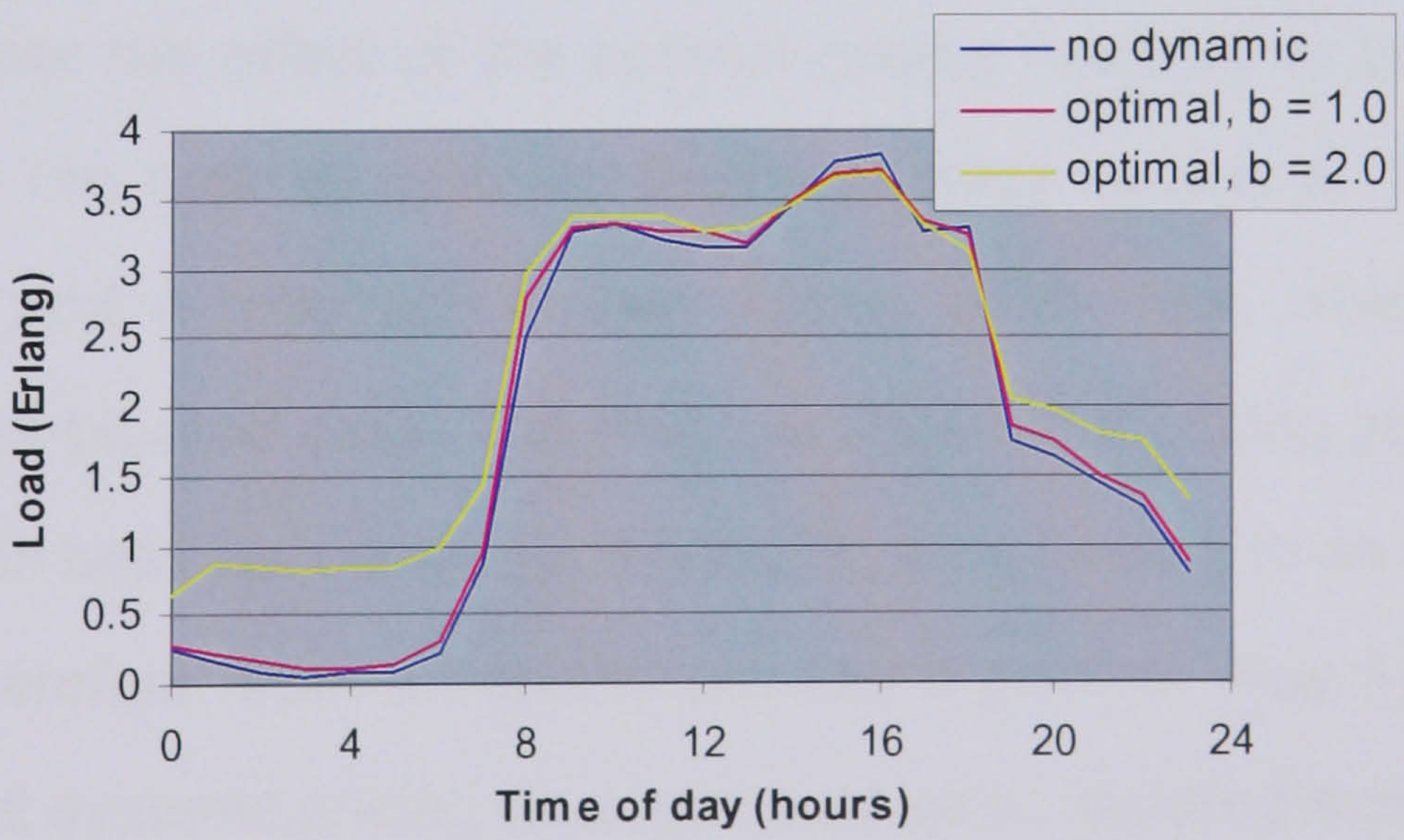

Figure F-26 *Load in the network as a function of time with optimal dynamic pricing for different demand elasticity*

As with the other pricing functions as β increases, its effect on the distribution of the load becomes more significant, with an increase in the number of off-peak calls and a decrease in the number of peak-hour calls.

In Figure F-27 the effect on network load of the optimal pricing is compared to the revenue attainment pricing for unit elastic demand ($\beta = 2.0$).

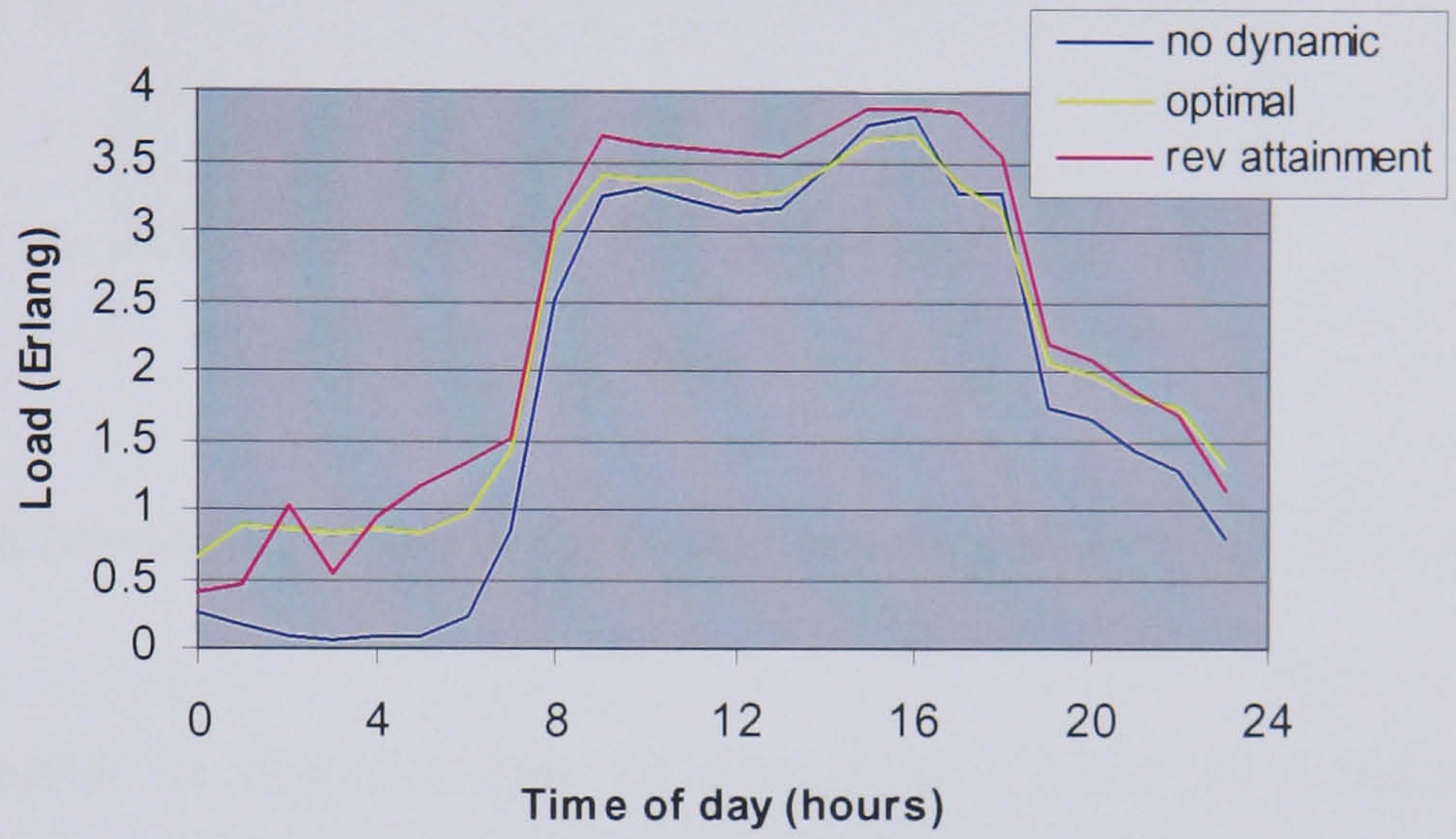

Figure F-27 *Load in the network as a function of time,* $\beta = 2.0$ *and optimal and ad hoc competition driven pricing.*

The change caused by the revenue attainment pricing function is more significant than the effect of the optimal pricing function, in particular during peak hours. The optimal dynamic pricing strategy causes an increase in off-peak hours and a reduction in load during peak hour, which controls the percentage of blocked calls. The revenue attainment pricing strategy leads to an increase in both peak and off-peak traffic, thus leading to an increase in call blocking. Therefore, from a network provider's point of view the performance of the optimal dynamic pricing strategy could be considered more satisfactory.

Running the simulation also indicated that with the introduction of optimal dynamic prices the total number of serviced calls would increase by 3% and 18% for $\beta = 1.0$ and $\beta = 2.0$ respectively (see Figure F-28). This can be attributed to the reduction in the average price in the network (by 12% for

$\beta = 1.0$ and 15% for $\beta = 2.0$) with the optimal dynamic pricing strategy (Figure F-29)

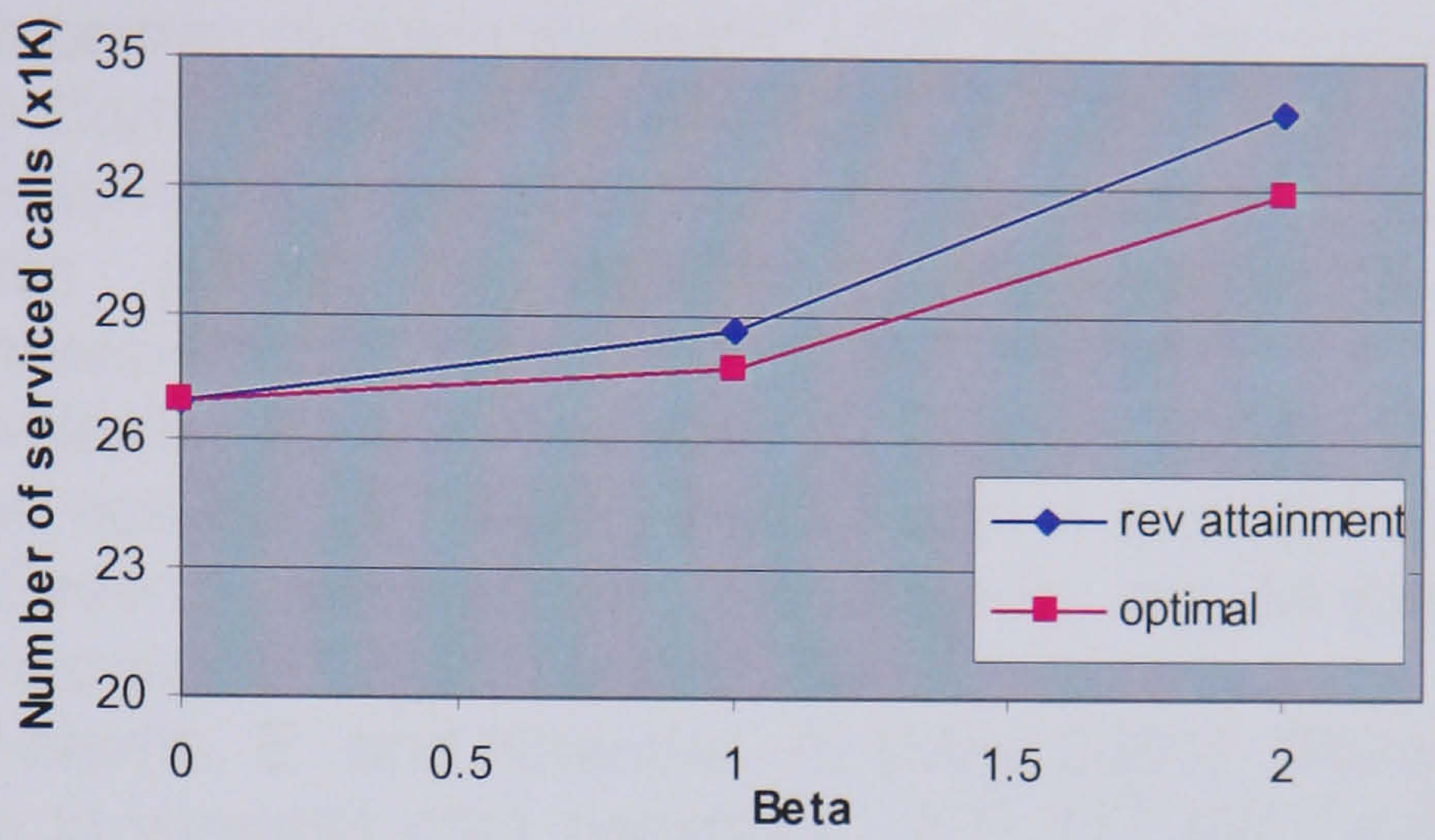

Figure F-28 *Number of successful calls with optimal and revenue attainment pricing*

This increase in the number of successful calls is smaller than the increase in calls generated by the alternative revenue attainment pricing strategy (by 3% and 5% for inelastic and unit elastic demand respectively). This is due to the even lower (by 35%) average price with the revenue attainment pricing (Figure F-29), which over-stimulates demand thus increasing both the number of successful and the percentage of blocked calls in the network.

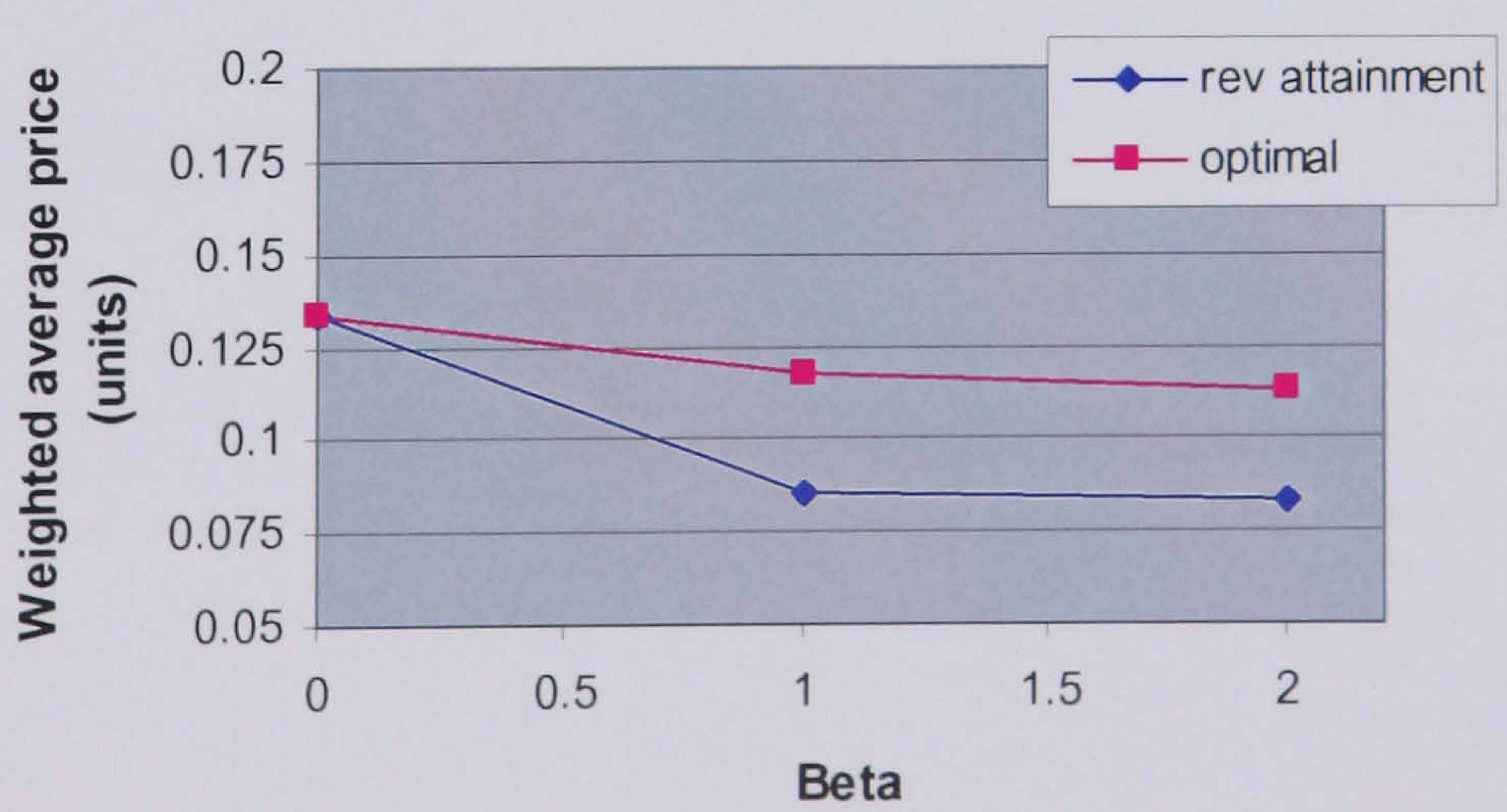

Figure F-29 *Weighted average network price with optimal and revenue attainment pricing*

Appendix G – List of Publications.

1. Fitkov-Norris, E. and Khanifar, A. (March 2000), "Dynamic pricing in mobile communication systems", *IEE First International Conference on Mobile Communication Technologies 3G2000*, London, 416-420.
2. Fitkov-Norris, E. and Khanifar, A. (September 2000), "Evaluation of dynamic pricing in mobile communication systems", *London Communications Symposium (LCS)*, London.
3. Fitkov-Norris, E. and Khanifar, A. (March 2001), "Dynamic pricing in cellular network, a mobility model with a provider-oriented approach", *IEE Second International Conference on Mobile Communication Technologies 3G2001*, London, 63-67.
4. Fitkov-Norris, E. and Khanifar, A. (May 2001), "Relevance of dynamic pricing in wireless data networks", *IEE 17th UK Teletraffic Symposium on Networks – A Service Centred Approach*, Dublin, Ireland.
5. Fitkov-Norris, E. and Khanifar, A. (September 2001), "Optimal price setting in dynamically priced mobile cellular networks", *IEEE International Telecommunications Symposium*, Tehran, Iran.
6. Fitkov-Norris, E. and Khanifar, A. (June 2002), "Effectiveness of dynamic pricing in mobile cellular networks for revenue generation and market share retention" 28th International Summer School: Applications of Mathematics in Engineering and Economics, Sozopol, Bulgaria.

References

[1] http://www.de.infowin.org/ACTS on 26.08.1999

[2] Bond, C., Wilkins, D. and Ablett, S. (1999) Will wireless win?, Analysys Publications, 1999.

[3] (1999), Proposal by the Commission for a Decision of the European Parliament and of the Council on the co-ordinated introduction of mobile and wireless communications (UMTS) in the Community, available at: http://www.de.infowin.org/ACTS/ANALYSYS/CONCENTRATION/MOBILITY on 26/08/99.

[4] MSI Data report: Telecommunication Services UK, April 1999.

[5] Ramsdale, P. (2000) UMTS Deployment Issues, UCL Lecture Notes.

[6] Arthur, C. (25th November 1997) Britain wired for fun, not wisdom, The Independent.

[7] Lam. D., Cox, D. and Widom, J. (1997) Teletraffic modelling for personal communications services, IEEE Communications Magazine, vol. **35** (2), 79-87.

[8] Kelly, T. (1996) Forecasting the Mobile Communications Market: A Finger in the Airwaves?, IRR Conference "Market Forecasting in the Telecoms Industry", Hong Kong.

[9] Boden, T. (1996) Market Forecasting in the Telecoms Industry, IRR Conference "Market Forecasting in the Telecoms Industry", Hong Kong

[10] Garg, V. and Wilkes J. (1996) Wireless and Personal Communications Systems, Prentice Hall.

[11] Garrard, G. (1998) Cellular Communications: Worldwide Market Development, Artech House Publishers.

[12] Macario, R. (1997) Cellular Radio Principles and Design, Macmillan Press Ltd.

[13] Aghvami, H. (2000), 3G200 IEE International Conference, Tutorial 3 UMTS/IMT2000.

[14] Gallagher, M. and Webb, W. (March 1999) UMTS the next generation for mobile radio, IEE Review, 59-63.

[15] Silva, S. da *et al.* (1997) Evolution towards UMTS at: http://www.infowin.org/ACTS/IENM/CONCENTRATION/MOBILTY/umts0.htmEuropean Commission, on 19/12/97.

[16] Steele, R. (1992) Mobile Radio Communications, Pentech Press.

[17] Mouly, M and Pautet M. (1992) The GSM System for Mobile Communications, Cell & Sys Publishers.

[18] EN 301 344 v6.3.2 (1999-07) General Packet Radio Service (GPRS); Service description; Stage 2 (GSM 03.60 version 6.3.2 Release 1997).

[19] ETSI TS 101 350 V7.0.0 (1999-07), Overall description of the GPRS radio interface; Stage 2 (GSM 03.64 version 7.0.0 Release 1998).

[20] http://www.totaltele.com.

[21] Universal Mobile Telecommunication System (UMTS); UTRAN Overall Description (3G TS 25.401 version 3.3.0 Release 1999).

[22] http://www/radio.gov.uk/ra_info/ra365.html

[23] Aghvami, H. Introduction to Digital Mobile Radio Communication Systems, Lecture Notes, University College London, 1998.

[24] Viterbi, A., CDMA Principles of Spread Spectrum Communication, Addison-Wesley Pubs.

[25] Concept Group Alpha – Wideband Direct-Sequence CDMA: System Description Summary, ETSI SMG #24, TR 101 146 v3.0.0 (1997-12).

[26] Black, U. (1996) Mobile and Wireless Networks, Prentice Hall Publishers.

[27] Hong, D. and Rappaport, S. (1986) Traffic model and performance analysis for cellular mobile radio telephone systems with prioritised and non-prioritised handoff procedures, IEEE Transactions on Vehicular Technology, vol. **VT- 35**(3), 77-92.

[28] Spilling, A., Nix, A., Beach M. (2000) Adaptive cell sizing in cellular networks, IEE Colloquium Capacity and Range Enhancement Techniques for the Third Generation Mobile Communications and Beyond.

[29] Gibson, J. ed. (1996) The Mobile Communications Handbook, CRC Press & IEEE Press.

[30] Elliot, S. and Dailey, D. (1995) Wireless Communications for Intelligent Transportation Systems, Artech House Publishers.

[31] What Phone? (June 1999) The Carphone Warehouse Communications magazine, pp. 38

[32] Boucher, J. (1988) Voice Teletraffic Systems Engineering, Artech House Publishers.

[33] Glad, T. and Ljung L. (2000) Control Theory Multivariable and Nonlinear Methods, Taylor and Francis Pubs.

[34] Kotler, P. (1988) Marketing Management – Analysis, Planning, Implementation and Control, Prentice Hall International Pubs.

[35] Ward, K. (1999) BT MSc Module 14 Lecture Notes, University College London.

[36] Ovum report on Mass Marketing Mobile Telephony (1998), at http://www.inellact.nat.bt.com/intellact/reports/analysis/mmm on 26/04/99.

[37] Ovum report on Mobile Operator Benchmarking II (1998), available at: http://corporate.intra.bt.com/mobility/ovum/mao/cis/maocis07/oper_bch.htm on 07/10/99.

[38] Ovum report on United Kingdom Market (1998) available at: http://corporate.intra.bt.com/mobiliy/ovum/mao/cis/maosis10/uk.htm on 07/10/99

[39] Pikkarainen, H. Realising the 3G vision (2002) available at: http://europe.synopsys.com/dialog/euro_compiler/issue17/expert1.html on 11/03/02.

[40] GSM 2000 Survey – The future of the mobile industry, available at http://www.emc-database.com/website.nsf/index/gsm2000 on 28/04/00.

[41] Brittan, D. (July 1997) Reporter: spending more and enjoying it less? at http://www.techreview.com/articles/july97/brittan.html on 02/12/98.

[42] Ovum report on Service Innovation (1998), available at: http://corporate.intra.bt.com/mobility/ovum/mao/cis/maosic01/ser_pre.htm on 07/10/99.

[43] http://www.btcellnet.net on 15th August 2000.

[44] http://www.nttdocomo.com/ser.html on 8th August 2000.

[45] Vincent, G. (2001) Learning from i-mode, IEE Review, 13-18.

[46] http://www.totaltele.com on 10th August 2000.

[47] Fishburn, P., Odlyzko, A. and Siders, R. (1997) Fixed fee versus unit pricing for information goods: competition, equilibria, and price wars, First Monday at: http://www.firstmonday.dk/issues/issue2_7/odlyzko/index.html on 30/11/98.

[48] MacKie-Mason, J. and Varian, H. (1995) Pricing contestable network resources, IEEE Journal on Selected Areas in Communications, vol. **13**, 1141-1149.

[49] Mason, R. (2000) Simple Competitive Internet Pricing, Philosophical Transactions of the Royal Society: Mathematical, Physical and Engineering Sciences, vol. **358** (1773), 2309-2318.

[50] Kaussar, N., Briscoe, B. and Crowcroft, J. , (1999) A charging Model for Sessions on the Internet, IEEE Computer Society Proceedings on International Symposium on Computers and Communications.

[51] Mitchell, B. and Vogelsang, I. (1991) Telecommunications Pricing, Cambridge University Press.

[52] Stiller, B. (1998) Accounting, Charging, and Pricing of Communication Services – An Internet-based Exemplified Study of Influences, IEEE Globecom'98, vol. **2**, 1196-1201.

[53] Ferrari, D and Delgrossi, L. (1999) Charging for QoS, 6th International Workshop on Quality of Service.

[54] Cruickshank, D. (1998) Prices of Calls to Mobile Phones, A Statement Issued buy the Director General of Telecommunications.

[55] Breker, P and Williamson, C. (1996), A simulation Study of Usage-Based Pricing Strategies for Packet-Switched Networks, Proceedings of the 21st IEEE Conference on Local Computer Networks.

[56] Cocchi *et al.* (1993) Pricing in computer networks: Motivation, Formulation and Example, IEEE/ACM Transactions on Networking, vol. **1**(6), 614-627.

[57] Bohn, R. *et al.* (March 1994) Mitigating the coming internet crunch: multiple service levels via precedence, Technical Report, University of California.

[58] Clark, D. (March 1995) A model for cost allocation and pricing in the Internet, MIT Workshop on Internet Economics, at: http://www.press.umich.edu/jep/works/ClarkModel.html on 28/04/00.

[59] Songhurst, D and Kelly, F (1997) "Charging Schemes For Multiservice Networks", Proceedings of the 15th ITC, June.

[60] Odlyzko, A. (1999) Paris Metro Pricing: The minimalist differentiated services solution, 7th International Workshop on Quality of Service, London.

[61] Saraydar, C., Mandayam, N and Goodman, D. (2000) Power Control in a Multicell CDMA Data System Using Pricing, 52nd IEEE Vehicular Technology Conference, vol. **2**, 484-491.

[62] Zarnikau, Z., Baughman, M. and Mentrup, G. (1990) Design of electric rates: matching cost with price, Forum for Applied Research and Public Policy, vol. **5** (4), 5 –11.

[63] Herriges, J. *et al.* (1993) The response of industrial customers to electric rates based upon dynamic marginal costs, The Review of Economics and Statistics, vol. **75**, 446 – 454.

[64] Dawn, T. (1994) Load management – is it worth the trouble? Electrical Review, vol. **227**(20), 38-40.

[65] Phelps, A. and Allera, S. (October 1992) A study of real time pricing in the UK: the Midlands electricity experience, Proceedings of the 1992 International Energy and DSM Conference, Toronto, 777-789.

[66] Aubin, C. *et al.* (1995) Real - time pricing of electricity for residential customers: econometric analysis of an experiment, Journal of Applied Econometrics, vol.**10**, S171- S191.

[67] Belobaba, P. (1989) Application of a probabilistic decision model to airline seat inventory control, Operations Research, vol. **37**, No 2, 183-197.

[68] Brumelle, S. *et al.* (August 1990) Allocation of airline sets between stochastically dependent demands, Transportation Science, vol. **24**(3), 183-192.

[69] Smith, B., Leimkuhler, J. and Darrow, R. (1992) Yield management at American airlines, Interfaces, vol. **22**(1), 8-31.

[70] Koschat, M., Uhler, L. and Spinagesh, P. (1995) Efficient price and capacity choices under uncertain demand: an empirical analysis, Journal of Regulatory Economics, vol. **7**, 5 -26.

[71] EN 301 113 v6.2.1 (1999-08) General Packet Radio Service (GPRS); Service description; Stage 1 (GSM 02.60 version 6.2.1 Release 1997).

[72] Cai, J. and Goodman, D. (October 1997) General Packet Radio Service in GSM, IEEE Communications Magazine, 122-131.

[73] Ferrer, C. and Oliver, M. (1998) Overview and Capacity of the GPRS (General Packet Radio Service), IEEE Globecom'98, 106-110.

[74] Oliver, H. and Songhurst, D. (2001) Market-Managed Multi-Service Internet, Journal of IBTE, vol. **2**, part 1, 42-46.

[75] ETSI TS 101393 v7.4.0 (2000-02) General Packet Radio Service (GPRS); GPRS Charging (GSM 12.15 version 7.4.0 Release 1998).

[76] A. Khanifar (1996) Dynamic Charging in Mobile Communication Systems, BT Internal Report.

[77] Cook, M. and Farquhason, C. (1998), Business Economics: Strategies and Applications, London, Pitman Publishers.

[78] Cosgrove, J. and Linhart, B. (1979) Customer choices under local measured telephone service, Public Utilities Fortnightly, vol. **30**, 27-31.

[79] Varian, H. (1996) Intermediate Microeconomics A Modern Approach, W.W. Norton & Company.

[80] Koschat, M., Uhler, L. and Spinagesh, P. (1995) Efficient price and capacity choices under uncertain demand: an empirical analysis, Journal of Regulatory Economics, vol. **7**, 5 -26.

[81] Wilson, A. (1967) A statistical theory of spatial distribution models, Transportation Research, vol. **1**, 253-269.

[82] Evans, A. (1971) The calibration of trip distribution models with exponential or similar cost functions, Transportation Research, vol. **5**, 15-38.

[83] Bouchard, R. and Pyers, C. (1965) Use of gravity model for describing urban travel, Highway Research Records, vol. **88**, 1-43.

[84] Gibbens R. and Kelly, F. (1999) Resource pricing and the evolution of congestion control, available at http://www.statslab.cam.ac.uk/frank/evol.html on 13/10/99, to be published in Automatica, vol. **35**.

[85] Murphy J. and Murphy, L. (1995) Bandwidth Allocation by Pricing ATM Networks at: http://thorung.eeng.dcu.ei/~murphy/publ/band-price/band-rice.html on 4/12/97.

[86] Anania, L. and Solomon, J. (March 1995) Flat: the minimalist B-ISDN rate, MIT Workshop on Internet Economics, at: http://www.press.umich.edu/jep/works/AnaniaFlat.html on 28/04/00.

[87] Karsten, M, Schmitt, J., Wolf, L. and Steinmetz, R. (1998) Provider-Oriented Linear Price Calculation for Integrated Services, 7th International Workshop on Quality of Service, 174-183.

[88] Wang, Q., Peha, J. and Sirbu, M. (1996) Optimal pricing for integrated-services networks with guaranteed quality of service, Internet Economics, MIT Press, Bailey, J. and McKnight, L. Eds.

[89] MacKie-Mason J and Varian, H. (1995) Pricing Congestible Network Resources, IEEE Journal of Selected Areas in Communication, vol. **13**, no. 7, 1141-1149.

[90] MacKie-Mason, J. and Varian, H. (1995) Pricing the Internet, Public Access to the Internet, Kahin, B. and Keller, J. eds. Prentice Hall.

[91] Rayes, A. and Min, P. (1995) Capacity Expansion of Least Busy Alternate Routing with Shadow Price, IEEE Globecom'95, vol. **2**, 1369-1373.

[92] Lau, W., Erramilly, A., Wang, J. and Willinger, W. (1995) Self-Similar Traffic Parameter Estimation: A Semi-Parametric Periodogram-Based Algorithm, IEEE Globecom' 95, vol. **3**, 2225-2231.

[93] Hawa, M. (1999) A Study of Deterministic Trends in Self-Similar Web-Cache Traffic Using Wavelet Analysis, M. Sc. Dissertation, University College London.

[94] Gupta, A., Stahl, D. and Whinston, A. (March 1995) A priority pricing approach to manage multi-service class networks in real time, MIT Workshop on Internet Economics, at: http://www.press.umich.edu/jep/works/GuptaPrio.html on 28/04/00.

[95] Chu, K. and Altmann, J. (2000) Demand for Different Qualities of Service for Internet Access: A Review of INDEX Findings, Philosophical

Transactions of the Royal Society – Mathematical, Physical and Engineering Sciences, vol. **358**, no.1773, 2319-2334.

[96] Mitchell, B. and Vogelsang, I. (1991) Telecommunications Pricing, Cambridge University Press.

[97] Wan, F. Y. M. (1995) Introduction to the Calculus of Variations and its Applications, Chapman & Hall.

[98] Butkov, E. (1973) Mathematical Physics, Addison-Wesley Pubs.

[99] Irving, J. and Mullineux N. (1959) Mathematics in Physics and Engineering, Academic Press Pubs.

[100] Release 2.4 MIL 3, Inc. (1994), http://www.mil3.com.

[101] (1979), Bell Technical Journal, vol. **58** (1).

[102] Lee, W. (1995) Mobile Cellular Communications, McGraw Hill Pubs.

[103] Frost, V. and Melamed, B. (March 1994) Traffic modelling for telecommunications, IEEE Communications Magazine, 70-81.

[104] Phillips, C. and Harbor, R., (1991), Feedback Control Systems, Prentice-Hall International.

[105] White, P. (1982) Optimisation Over Time, John Wiley & Sons.

[106] Nash, S. and Sofer, A., (1996), Linear and Non-linear Programming, McGraw-Hill International.

[107] Gill, P. *et. al.* , (1981), Practical Optimisation, Academic Press.

[108] Pinch, E., (1993), Optimal Control and the Calculus of Variations, Oxford Science Publications.

[109] European Mobile Statistics 2002, at http://home.intekom.com/cellular/stats/stats-europe.htm on 30/07/2002.

[110] SPSS for Windows (1998) Release 8.0.2, Chicago, SPSS Inc.